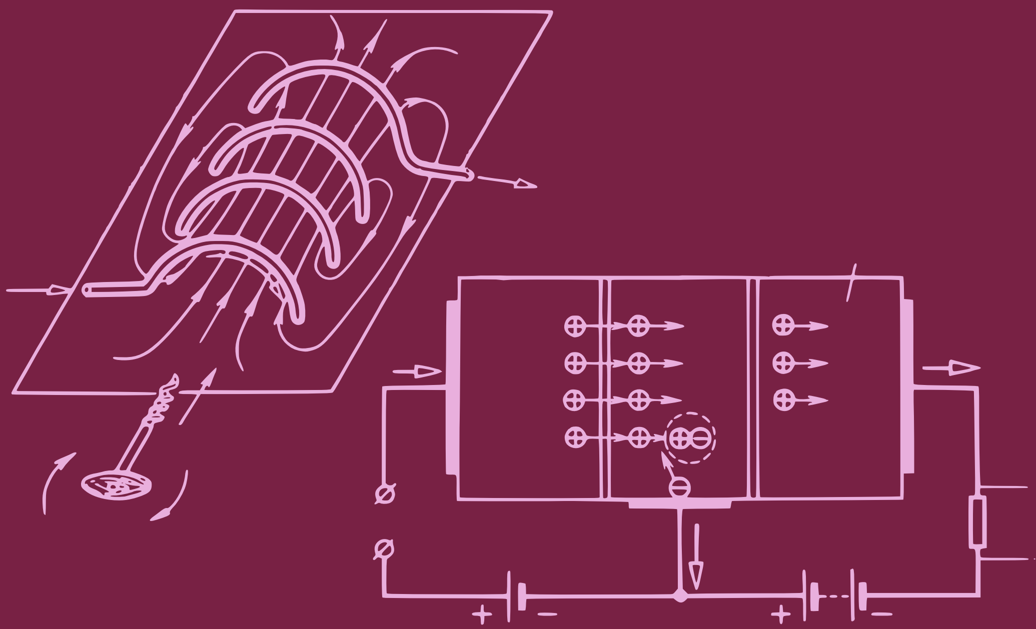
V.S. Popov, S.A. Nikolayev

Basic Electricity and Electronics



Mir Publishers • Moscow

ИЗДАТЕЛЬСТВО · МИР ·

В. С. Попов, С. А. Николаев

ОБЩАЯ ЭЛЕКТРОТЕХНИКА С ОСНОВАМИ ЭЛЕКТРОНИКИ

Энергия Москва 1977

V. S. Popov
and
S. A. Nikolaev

Basic Electricity and Electronics

Translated from the Russian

by

Alexander Kuznetsov

Mir Publishers • Moscow

First published 1979
Revised from the 1977 Russian edition

The Russian Alphabet and Transliteration

А а	a	К к	k	Х х	kh
Б б	b	Л л	l	Ц ц	ts
В в	v	М м	m	Ч ч	ch
Г г	g	Н н	n	Ш ш	sh
Д д	d	О о	o	Щ щ	shch
Е е	e	П п	p	Ъ ъ	″
Ё ё	e	Р р	r	Ы ы	y
Ж ж	zh	С с	s	Ь ь	′
З з	z	Т т	t	Э э	e
И и	i	У у	u	Ю ю	yu
Й й	y	Ф ф	f	Я я	ya

The Greek Alphabet

Α α	Alpha	Ι ι	Iota	Ρ ρ	Rho
Β β	Beta	Κ κ	Kappa	Σ σ	Sigma
Γ γ	Gamma	Λ λ	Lambda	Τ τ	Tau
Δ δ	Delta	Μ μ	Mu	Υ υ	Upsilon
Ε ε	Epsilon	Ν ν	Nu	Φ φ	Phi
Ζ ζ	Zeta	Ξ ξ	Xi	Χ χ	Chi
Η η	Eta	Ο ο	Omicron	Ψ ψ	Psi
Θ ϑ θ	Theta	Π π	Pi	Ω ω	Omega

На английском языке

Contents

Introduction

Electrical engineering has to do with practical applications of electric energy. This includes electric power generation, distribution, and conversion. Electric energy has very valuable properties: it can easily be derived from other forms (mechanical, chemical, etc.), transmitted with low losses for hundreds or even thousands of kilometers to homes and plants, distributed among users and converted back into mechanical, thermal, chemical, and other necessary forms. Electricity makes it possible to utilize the inexpensive energy accumulated by Nature (energy of falling water), or to cut down its cost (such as when generated by burning peat or low-grade coal).

Used on a large scale in all fields of national economy and everyday life, electricity promotes the introduction of advanced machinery and complex process mechanization and automation into industry. It has given life to new industrial processes, such as electric welding, electrolysis, and hardening by high-frequency currents. Owing to its abundance and low cost, electricity has provided novel approaches to many problems of industrial production and enabled many breakthroughs in science to become everyday practice and to raise labour productivity.

Electronics, which has now become a division of electrical engineering in its own right, considers the principles of operation, design and application of semiconductor, vacuum and gas-filled devices in science, various industries, and technology. For example, semiconductor and gas-filled rectifying devices are used in power engineering to convert alternating current to direct current for electric drives,

electric traction, electrochemical and other production processes.

Without semiconductor, vacuum and gas-filled devices it would be impossible to effect process automation, that is, to regulate and control production processes.

Rapid advances in computer engineering have made it possible not only to raise the performance of automatic control systems to a new level, but also to tackle nation-wide economic problems.

Electric and electronic devices for the generation, processing, transmission and display of data are key elements of automated management and control systems at any level.

Technologies have been developed by which a great number of circuit components (diodes, transistors, resistors, capacitors, etc.) can now be made in the form of film microcircuits and assembled into sophisticated systems. In the manufacture of these microcircuits, use is made of electron-beam and laser equipment. Electronics has found uses in the manufacture of superhigh-purity materials, such as tungsten, molybdenum, tantalum and niobium, essential to modern technology. It is obvious that a firm knowledge of the fundamentals of applied sciences, notably electrical engineering and electronics, is essential to any one who wishes to gain insight into present-day technology.

Part One Electricity

Chapter One Electric Field

1-1. General

Any body is made up of elementary particles, each carrying an *electric charge**. For example, a proton carries a positive charge and an electron, a negative charge. Some charged elementary particles make up atoms and molecules of substances, others are in the free state. A charged body is one in which positive charge prevails over negative or vice versa; an electrically neutral body has an equal number of negative and positive charges.

Moving elementary particles carrying electric charge, or simply electric charges are inseparable from the surrounding *electromagnetic field* which is a form of matter. The electromagnetic field consists of two interrelated components, an *electric field* and a *magnetic field*. Their existence is revealed by the action they produce on charged elementary particles or bodies.

Unlike charges attract and like charges repel one another. Since a charge cannot be isolated from the surrounding magnetic field, charged bodies interact via the electric field.

As the electric field exerts a force on any electrically charged body or particle placed in it, it is capable of doing work. Therefore, the electric field possesses what is called *electric energy*.

Electrically charged particles of a substance and their electric field are two inseparably linked forms of matter.

* Electric charge is a property of material particles or bodies, which characterizes their interaction with their own and an external electromagnetic field. As electric charge is a property of material particles or bodies, it cannot be divorced from matter; yet, when considering electromagnetic effects, it is customary to use "charge" in the sense of charged particles or bodies. In a quantitative definition, charge is synonimous with "quantity of electricity".

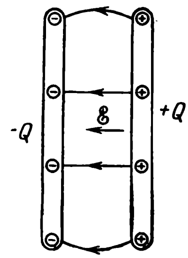


Fig. 1-1. Electric field between two parallel plates carrying dissimilar charges

Any point in an electric field can be characterized in terms of *electric field strength*, $\mathscr{E}$.

Electric field strength is the ratio of the force F that the field exerts on the point test charge Q placed at the point of interest, to this charge, that is,

$$\mathscr{E} = F/Q \tag{1-1}$$

A *point test charge* is a charged body very small in size, whose charge is so negligible that it does not practically distort the field being considered.

If Q is taken equal to unity (one coulomb), $\mathscr{E}$ is numerically equal to F, so the *electric field strength at a point is numerically equal to the force per unit charge*.

The electric field strength is characterized not only by its magnitude, but also by the direction in which the field exerts its force on a stationary positive charge at a given point. Thus, the *electric field is a vector*.

Figure 1-1 illustrates the vector of an electric field between two parallel plates with charges $+Q$ and $-Q$.

The electric field is graphically represented by *electric lines of force*. An electric line of force should be drawn in such a way that at any point the electric field vector is directed along a line tangent to it at that point.

An electric line of force begins at a positive charge and ends at a negative one, so it is an open line.

If we take a unit area (for example, 1 cm^2) normal to the vector and draw so many lines through it that their number is equal or proportional to the electric field strength in this area, the density of lines will characterize the magnitude of the electric field strength.

A field is said to be *uniform*, if its vector is the same at any point in space. An example is the electric field between parallel plates (Fig. 1-1), in an area sufficiently distant from the edges.

1-2. Voltage. Potential

Suppose that a positive test charge Q has been moved by a uniform field from point M to point N spaced a distance l (Fig. 1-2) apart.

The work done by the field on the charge, or the potential energy expended, is:

$$W = Fl$$

or, in view of Eq. (1-1),

$$W = Fl = \mathscr{E}Ql \qquad (1\text{-}2)$$

The quantity defined by the ratio of the work done in moving a charge Q between two points in the field to the charge is called the *voltage* between points M and N. So,

$$V = W/Q$$

Thus, the voltage between two points *is numerically equal to the work done by the field in moving a unit positive charge between these two points.*

Since $W = \mathscr{E}Ql$, we may write

$$V = W/Q = \mathscr{E}Ql/Q = \mathscr{E}l \qquad (1\text{-}3)$$

In the International System of Units, the unit of length is the metre (m); the unit of mass is the kilogram (kg); the unit of time is the second (s); the unit of force is the newton (N); the unit of work is the joule (J); the unit of electric charge is the coulomb (C); the unit of voltage is the volt (V).

According to Eq. (1-3):

$$1\ \mathrm{V} = 1\ \mathrm{J}/1\ \mathrm{C}$$

By Eq. (1-3), the electric field strength may be defined as

$$\mathscr{E} = V/l \qquad (1\text{-}4)$$

Hence, the electric field strength has the dimension of volts per metre:

$$[\mathscr{E}] = \mathrm{V/m}$$

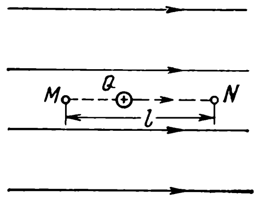


Fig. 1-2. Motion of an electric charge $+Q$ in a uniform field

The voltage between a given point M in an electric field and any other point in the same field, whose potential is assumed equal to zero, is called the *potential* (φ) of the given point. The potential of the earth's surface is frequently taken as zero. The potential of a point in a field *is numerically equal to the work that the field would have to do in moving a unit positive charge from that point to the point whose potential is taken as zero.*

When a unit positive charge is moved from point M with a potential φ_M to point N with a potential φ_N, the work done by the field, that is, the voltage between these points, is equal to the potential difference:

$$V_{MN} = \varphi_M - \varphi_N \tag{1-5}$$

or in words, *the voltage between two points in an electric field is equal to the potential difference between them.* Potential is measured in volts, the same units as voltage.

If all points on a surface have the same potential, the surface is called equipotential.

1-3. Electric Conduction

In the chemical elements that make up any substance, every atom consists of a positively charged nucleus around which are distributed negatively charged electrons. Atoms are usually electrically neutral, because the charge on the nucleus is equal to the total charge on the electrons.

When a neutral atom or molecule loses an electron, it turns into a *positive ion.* The electron which has left the atom either attaches itself to another neutral atom to form a *negative ion* or remains free. Such free electrons are called *conduction electrons*, and the process by which ions are for-

ned is identified as *ionization*. The number of free electrons or ions per unit volume of a material is termed the *concentration* (or number density) *of charge carriers*, symbolized by the letter "n".

When a material is placed in an electric field, the field forces cause a directed flow of charge carriers (conduction electrons or ions), which is referred to as an *electric current*. The property of a material to produce an electric current under the action of an electric field is called *electric conduction*. The extent of this property is called the *conductivity* of that material (see Sec. 2-4). The electric conductivity of a material depends on the concentration of charge carriers and increases with concentration. According to their conductivity and the manner in which some physical factors affect it, all materials can be divided into *conductors*, *dielectrics* (insulating materials) and *semiconductors*.

Conductors have high conductivity; among them are most metals and their alloys, carbon, electrolytes (aqueous solutions of salts, acids and alkalis), and melts.

In contrast, dielectrics possess insignificant conductivity; among them are gases, mineral oils, varnishes and a great number of non-metallic substances.

Semiconductors have conductivity which is between those of metals and dielectrics. They include such elements as silicon, germanium and selenium, or oxides of metals. A typical feature of semiconductors is that their conductivity strongly depends on external factors.

Each electron in an atom can take only a discrete set of values for its energy, that is, its energy can change in certain small portions called *quanta* and it may only occupy certain permitted or allowed energy states or levels. The transition of an electron to a higher energy level, that is, to a more distant orbit, requires some energy to be spent to overcome the force of attraction by the nucleus. Thus, the electrons more distant from the nucleus have a greater amount of energy. The transition of an electron to a lower level is accompanied by the emission of the energy given up by the atom.

In solids which consist of a multitude of atoms, the energy levels, due to the mutual influence of adjacent atoms, fall into groups, or energy bands. These bands are separated by

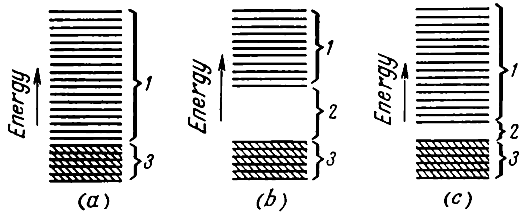


Fig. 1-3. Energy levels
(a) conductor; (b) dielectric; (c) semiconductor; 1—allowed (empty) band; 2—forbidden band; 3—filled band

areas where there can be no electrons and are known as *forbidden bands* or energy gaps.

The energy bands corresponding to the permitted levels can be divided into a filled band and an empty band. In order to provide electric conduction, some electrons must leave the filled band and move to the empty band. The possibility of such a transfer is determined by the width of the forbidden band, and the energy that must be spent for this transfer is in proportion to the width of the band.

The difference in electric conductivity between conductors, semiconductors and dielectrics can be explained by differences in their structure. According to the band theory of solids, metal conductors have high electric conductivity due to the fact that their filled bands closely adjoin the empty bands (Fig. 1-3*a*). In this case, the electrons in metals are free to move from the filled to the empty band. So in metal conductors, the electrons can easily be made free. This easily obtainable considerable concentration of electrons ensures high electric conductivity for conductors.

When a voltage is applied to a conductor, it sets up an electric field. This field turns the random motion of free electrons into a regular one, and they flow in the direction opposing the direction of the field (because they are negative), so an electric current begins to flow in the conductor.

When the filled and empty bands of a material are separated by a sufficiently broad energy gap (Fig. 1-3*b*), the latter does not practically allow electrons to move into the empty band. Thus, both the concentration of free electrons and conductivity of the material are negligible, so this material is a dielectric.

In semiconductors, the energy gap is much narrower than it is in dielectrics (Fig. 1-3*c*). Consequently, a small amount of excitation is sufficient to move electrons to the empty band; such an excitation can be provided, for example, by the build-up of thermal motion of atoms with a rise in temperature. Because of this, semiconductors possess conductivity which is between those of conductors and dielectrics and strongly depends on external factors, for example, temperature.

Conductors in which an electric current is carried only by the flow of electrons are said to possess *electron* (or electronic) *conduction*; they are also referred to as conductors of the first type. These are mainly metals and alloys.

Conductors in which an electric current is carried by the flow of positive and negative ions are said to have *ion* (or ionic) *conduction*; they are called conductors of the second type. Among them are electrolytes, that is, aqueous solutions of acids, salts and alkalis.

1-4. Capacitance. Capacitors

An *electric capacitor* is a system of two conductors (plates) separated by a dielectric.

Examples of natural capacitors are a pair of power-transmission line wires, two cable conductors, a cable conductor and its sheath, or a partition insulator (isolating a wire from a wall or a metal case). Various types of capacitors are widely used, in particular, capacitors formed by parallel metal plates isolated from each other.

The graphic conventions of capacitors are shown in Fig. 1-4.

The plates of a capacitor accumulate and store charges equal in value and opposite in sign. The electric charge Q on each plate is in proportion to the voltage V between the plates; this suggests that

$$Q = CV$$

The quantity C equal to the ratio of the charge on one plate to the voltage between the plates is called the *capacitance of a capacitor*. Mathematically, it is written

$$C = Q/V \tag{1-6}$$

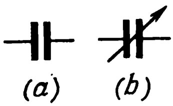


Fig. 1-4. Graphic conventions for capacitors
(a) fixed capacitor; (b) variable capacitor

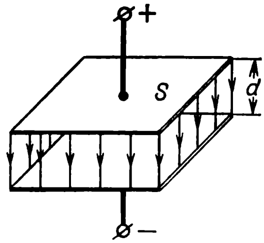


Fig. 1-5. Plane-parallel capacitor

In the International System, the unit of charge is the coulomb and the unit of voltage is the volt; so the unit of capacitance is the coulomb divided by the volt. It is called the *farad* (F):

$$1 \text{ farad} = 1 \text{ coulomb}/1 \text{ volt}$$

In practice, use is made of smaller units, such as the microfarad ($1\ \mu F = 10^{-6}$ F) or the picofarad ($1 \text{ pF} = 10^{-12}$ F).

The capacitance of a capacitor depends on the shape and size of its plates, their relative position, the spacing between them, and the properties of the dielectric separating the plates.

For example, the capacitance of a parallel-plate capacitor whose plates are placed in a vacuum (Fig. 1-5) is

$$C = \varepsilon_0 S/d$$

where S = area of the plates, m^2
d = spacing between the plates, m
ε_0 = electric constant characterizing the electric field in a vacuum

The dimension of the electric constant is

$$[\varepsilon_0] = \left[\frac{Cd}{S}\right] = \frac{\text{farad} \times \text{metre}}{\text{metre}^2} = \frac{\text{farad}}{\text{metre}} \ (\text{F/m})$$

The value of the electric constant depends on the system of units adopted. In the International System, it is

$$\varepsilon_0 = 10^{-9}/36\pi = (\text{approx.})\ 8.85 \times 10^{-12} \text{ F/m} \qquad (1\text{-}7)$$

The dielectric properties of various materials can be compared with the properties of a vacuum.

If the space between the plates of a parallel-plate capacitor is filled with a dielectric, its capacitance will rise ε_r times, and will be given by the following formula

$$C = \varepsilon_0 \varepsilon_r S/d = \varepsilon_a S/d \qquad (1\text{-}8)$$

The multiplier ε_r called *relative permittivity* (or dielectric constant), is a dimensionless number. The relative permittivities of some dielectrics are given in Table 1-1.

Table 1-1

Characteristics of Some Insulating Materials

Material	$\mathscr{E}_{bd}$	ε_r	ρ
	10^3 kV/m	–	ohms m
Oil-treated paper	10-25	3.6	—
Air	3	1	—
Paper-base laminate	10-15	4-7	10^8-10^{10}
Micanite	15-40	5-6	10^9-10^{11}
PVC	32.5	3.2	10^{12}
Rubber	15-20	3-6	10^{11}-10^{12}
Glass	10-15	6-10	10^{12}
Mica	50-100	5.4	5×10^{11}
Sovol	15	5.3	10^{11}-10^{12}
Transformer oil	5-18	2-2.5	5×10^{12}-5×10^{13}
Porcelain	15-20	5.5	10^{12}-10^{13}
Electric-grade pressboard	8-12	3-5	10^6-10^8

The product of the relative permittivity and electric constant of a medium gives its absolute permittivity:

$$\varepsilon_a = \varepsilon_r \varepsilon_0$$

Capacitors are made in a variety of sizes, that is, capacitances (1 pF to 1000 μF), with nominal voltages up to 100 kV, and differing in design and application.

Paper, mica and ceramic capacitors are used both in d.c. and a.c. circuits, whereas electrolytic capacitors are used only in d.c. circuits.

A paper capacitor (Fig. 1-6) consists of two long tapes of aluminium foil interleaved with paraffin-treated paper.

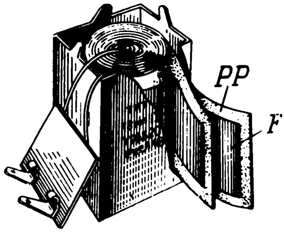


Fig. 1-6. Paper capacitor
F—foil; PP—paraffin-treated paper

Electrolytic capacitors employ the aluminium oxide film on one of the plates as the dielectric. The other plate is paper or cloth soaked with a thick solution of an electrolyte.

1-5. Connection of Capacitors

In order to obtain a necessary capacitance or when the line voltage exceeds the nominal voltage of a capacitor, capacitors can be connected in series, in parallel or in series-parallel.

In a series arrangement of capacitors (Fig. 1-7), each capacitor has the same charge because the charge is applied from the power source only to the outer plates, while the charges on the inner plates are produced upon the separation of charges which earlier cancelled one another.

Calling the charge of one capacitor Q, we may write for two capacitors connected in series

$$V_1 = Q/C_1 \quad \text{and} \quad V_2 = Q/C_2$$

that is, the voltage across the capacitors is not the same for different values of capacitance.

Expressing the voltage across the terminals of the circuit

$$V = V_1 + V_2$$

in terms of the ratio of the charges to the capacitance of the capacitors, we get:

$$Q/C = Q/C_1 + Q/C_2$$

or, eliminating Q, we obtain in the final form

$$1/C = 1/C_1 + 1/C_2 \tag{1-9}$$

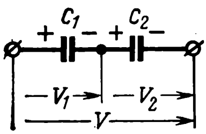


Fig. 1-7. Series connection of capacitors

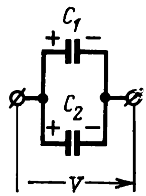


Fig. 1-8. Parallel connection of capacitors

Hence, the total or equivalent capacitance of two capacitors connected in series is

$$C = \frac{C_1 C_2}{C_1 + C_2} \tag{1-10}$$

In a parallel arrangement of capacitors (Fig. 1-8), the voltage across each capacitor is the same. The charge on each capacitor is

$$Q_1 = C_1 V \text{ and } Q_2 = C_2 V$$

The charge received by all capacitors connected in parallel is equal to the sum of the charges of the individual capacitors. For two capacitors connected in parallel

$$Q = Q_1 + Q_2$$

Hence, the total or equivalent capacitance is

$$C = Q/V = \frac{Q_1 + Q_2}{V} = C_1 + C_2 \tag{1-11}$$

that is, the sum of the capacitances of the individual capacitors.

Given any number of capacitors arranged in series or in parallel, it is easy to define their equivalent capacitances using Eqs. (1-9) and (1-11).

Example 1-1. Determine the equivalent capacitance of two capacitors connected in series and in parallel: $C_1 = 2\ \mu\text{F}$; $C_2 = 4\ \mu\text{F}$.

Solution.

For the series connection of the capacitors, the equivalent capacitance is

$$C=\frac{C_1C_2}{C_1+C_2}=\frac{2\times 4}{2+4}=1.33\ \mu\text{F}$$

In the case of the parallel connection, the equivalent capacitance is

$$C = C_1 + C_2 = 2 + 4 = 6\ \mu\text{F}$$

1-6. Energy of the Electric Field

When the voltage across a capacitor connected to a power source is raised, the charges on its plates and the electric field strength of the dielectric increase, too. Naturally, the electric field energy of the capacitor builds up as well, due to the power supplied from the source.

The increase in the voltage across the capacitor, dV, brings about an increase in the electric field energy, Eq. (1-3), of the capacitor

$$dW_C = QdV$$

The total amount of energy W_C accumulated in the elecr tric field of the capacitor during the rise in the voltage across its terminals from $v_C = 0$ to $v_C = V_C$ can be determined by summing the elementary energies dW_C. Thus, the electric field energy of the capacitor can be written

$$W_C=\int_{v_C=0}^{v_C=V_C} Q\,dv_C=C\int_0^{V_C} v_C\,dv_C=\frac{CV_C^2}{2}=\frac{QV_C}{2} \quad (1\text{-}12)$$

If we disconnect a charged capacitor from its power source and connect together (short-circuit) its plates through a conductor, the capacitor will discharge, and the short-duration discharge current will generate in the conductor an amount of heat which is equivalent to the potential field energy of the charged capacitor.

Example 1-2. Determine the energy accumulated in the electric field of a capacitor with a capacitance of 10 μF, if the voltage across the capacitor is 300 V.

Solution.

The energy of the electric field is

$$W_C = CV_C^2/2 = \frac{10 \times 10^{-6} \times 300^2}{2} = 0.45 \text{ J}$$

1-7. Polarization of Dielectrics

In a dielectric placed in an electric field, the field forces shift the electron orbits in a direction opposite to the field, so the atomic nuclei are no longer at the centre of the orbits (Fig. 1-9*a*) but some distance from it (Fig. 1-9*b*).

As far as electric properties are concerned, such an atom may be treated as an *electric dipole*, that is, a pair of opposite point charges separated by a small distance l (a dipole arm) from each other. The charges forming the dipoles of a dielectric are called coupled, and the product of the charge Q and the arm l is termed the *electric moment of a dipole*, or the *dipole moment*

$$p = Ql$$

The electric moment is a vector directed from the negative to the positive charge of a dipole unit.

Thus, when placed in an external electric field, atoms become dipoles, and their dipole moments p tend to orient themselves in the direction of the external field. When the electric field is removed, the shift of the electron orbits vanishes, too. This shift is called the *polarization of a dielectric*.

Polarized atoms and molecules set up an electric field of their own, which is opposite to the direction of the main field, thereby causing the main field to weaken. The polarizability of a dielectric under the action of an electric field is defined in terms of its permittivity which shows how many times the main field weakens due to polarization.

In a dielectric subjected to an alternating electric field, the shift is also alternating, and this causes the dielectric to heat.

The faster the electric field alternates, the more the dielectric heats. This effect (dielectric heating) is used to dry dielectrics or to carry on chemical reactions requiring elevated temperatures. The energy per unit volume spent to

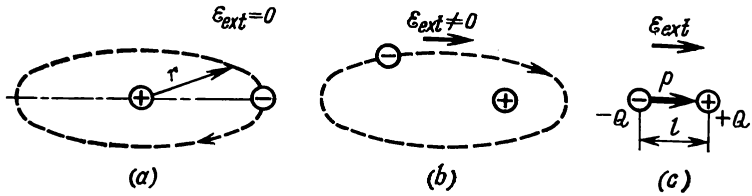


Fig. 1-9. Unpolarized molecule
(*a*) in the absence of an external field; (*b*) in the presence of an external field; (*c*) an equivalent dipole

heat a dielectric by an alternating shift of charges is termed *specific dielectric loss.*

If the applied electric field is too strong, the dielectric will *break down*, i.e., it will be locally destroyed. The respective value of electric field strength $\mathscr{E}_{bd}$ is called the *dielectric strength of the material*, and the respective voltage is termed the *breakdown voltage* V_{bd}.

The breakdown of a dielectric may take several forms. One of them is due to the fact that the applied electric field causes the few available electrons to reach a certain critical velocity sufficient to break away new electrons from the neutral atoms and molecules of the dielectric; this leads to *impact ionization* followed by breakdown. Another form is due to the heating of a dielectric in an electric field. In this case, the dielectric is damaged thermally, for example, cracks, chars, etc. The heating may be caused by dielectric loss, an increase in the conductivity of the dielectric, or a considerably disproportionate rise in volume current (see Sec. 2-4) due to an increase in voltage.

The dielectric strength of a material depends on several factors: voltage, rate of voltage change, time of exposure to voltage, form of the electric field (shape of electrodes), thickness of the material, its temperature, humidity, and pressure (in the case of gases).

For reliable operation of an electric device, it is important that all its dielectric parts shall not be exposed to electric field strengths exceeding a maximum safe (or allowable) value which should be a fraction of the dielectric strength. The dielectric strength of some materials is given in Table 1-1.

Example 1-3. A sheet of electric-grade pressboard 0.3 cm (3×10^{-3} m) thick is clamped between two flat metal electrodes. Determine the maximum safe and breakdown voltages. The maximum safe voltage should be one-third of the breakdown voltage.

Solution.

According to Table 1-1, the dielectric strength of pressboard is $\mathscr{E}_{bd} = 10 \times 10^3$ kV/m.

The breakdown voltage is

$$V_{bd} = \mathscr{E}_{bd} d = 10 \times 10^3 \times 3 \times 10^{-3} = 30 \text{ kV}$$

The maximum safe (or allowable) voltage is

$$V_s = V_{bd}/3 = 30/3 = 10 \text{ kV}$$

1-8. Electric Insulating Materials

An electric insulating material is that which has a negligible electric conductivity.

Electric insulating materials are used in electric devices to insulate conductors or parts operated or held at different potentials. They should have a sufficient dielectric strength (see Table 1-1), high volume and surface resistivity (see Sec. 2-4), high stability of electrical and mechanical characteristics, and low dielectric loss. Mechanical, thermal and other properties are also important.

This diversity in requirements has led to the use of a great number of different electric insulating materials which may be classified in many ways. For example, there are gaseous, liquid and solid insulating materials; organic and inorganic materials. Also, they are divided into many classes according to their heat resistance.

Gaseous Insulators. The principal gaseous insulator is air. At a normal temperature of +20°C and a normal pressure of 0.13 MPa (760 mm Hg), the dielectric strength of air (3×10^3 kV/m) is lower than that of most liquid and solid dielectrics. Therefore, the air gap breaks down sometimes near the surface of a dielectric, and this is called *surface breakdown*.

Hydrogen, carbon dioxide, nitrogen and inert gases are also used as insulators.

Liquid Insulators. Liquid insulators are mineral oils, synthetic liquids, resins and varnishes.

Mineral oils are obtained from petroleum by distillation and comprise mixtures of liquid hydrocarbons. They are used in oil transformers, oil circuit-breakers, cables and capacitors.

In transformers, oil is used to insulate current-conducting parts and cool them by convection, i.e., transfer heat by its circulation.

In oil circuit-breakers, oil additionally quenches the arc as the circuit-breaker is opened.

In cables and capacitors, it is used to treat paper insulation.

Oil should have a high dielectric strength (10 to 20 MV/m). As the strength drops sharply in the presence of moisture, oil should be dried and purified before use and regularly in service. Some characteristics of oil are given in Table 1-1.

Sovol, a Soviet-made artificial liquid insulator, is a mixture of variously chlorinated molecules of diphenyl. It is used instead of mineral oil to impregnate and fill capacitors. This treatment doubles their capacitance.

Transformers are also filled with Sovtol which is Sovol diluted with trichlorbenzene. As Sovtol does not burn, the transformers filled with it are fire-proof.

At low temperatures, resins are amorphous vitreous materials. When heated, they become first plastic, then liquid. Resins do not absorb water and do not dissolve in it, but they dissolve in alcohol and other solvents. Resins are the most important components of many varnishes, plastics and films.

Natural resins are obtained from excretions of certain insects (for example, shellac) or from the sap of trees. Most important are synthetic (polymeric) resins such as polyethylene and PVC (see Table 1-1). Plastics based on them are used as wire and cable insulation, protective coatings, and in the manufacture of varnishes.

Varnishes are solutions of resinous materials, such as gums, bitumens, drying vegetable oils (for example, linseed), and cellulose ethers. As they dry, varnishes form a hard coating. Varnishes serve to impregnate insulators to protect them against exposure to moisture and chemically active

media, and also to bond together sheets of mica, paper and fabric.

Solid Insulators. Solid insulators are the most numerous group of insulating materials (see Table 1-1).

1. Fibrous organic materials, such as paper, pressboard, fibre, cloth, are made of fibres of wood, cotton or capron.

They possess flexibility, sufficient mechanical strength and appreciable hygroscopicity (tendency to absorb water) which can be reduced by impregnation with mineral oil or compounds.

Paper is mainly manufactured from wood. There exist several grades of insulating paper: for cables, capacitors, bakelite sleeves, electrical-sheet steel insulation, etc.

Pressboard is made of cellulose under pressure. It is widely used for spacers in electric devices, transformers and other apparatus.

Vulcanized or hard fibre is produced by treating porous cotton-base paper with zinc chloride. Fibre is made into panels, mounts, bushings, etc.

Paper-base laminates consist of many sheets of paper pressed together and bonded with bakelite varnish.

2. Plastics are mixtures consisting essentially of binders and fillers. Binders are polymers (synthetic resins) and also liquid glass or cement.

Plastics are widely used in electrical engineering as insulating and structural materials.

3. Elastomers are materials possessing elasticity, that is, they readily elongate when stretched and restore their former size and shape when the load is removed.

Rubber (natural and synthetic) has high elasticity and low penetrance for moisture and gases.

Vulcanized rubber is an elastic material obtained by the addition of sulphur to rubber, followed by vulcanization at an elevated temperature. Vulcanized rubber containing 1 to 3 per cent sulphur is soft and elastic; a rubber compound containing 25 to 50 per cent sulphur is called hard rubber (or ebonite); it is a nonelastic material which can be worked and machined.

Nowadays, rubber is being replaced by elastic plastics such as PVC or polyethylene which can better resist attack by alkalis, acids and mineral oils.

4. Glass is made by fusing together silica (SiO_2) and oxides of sodium, potassium and calcium, followed by cooling. Ordinary glass is brittle. Special grades of glass are made break-proof.

In electrical engineering, glass is manufactured into insulators, envelopes for lamps and radio tubes.

Glass fibre and cloth are used to insulate wires intended for operation at high temperatures.

5. Electrical porcelain is made from kaolin (china clay), fireclay, quartz and feldspar. Porcelain products are dipped in glaze and then fired to convert the glaze into a glass coating which reduces the moisture absorption of porcelain.

Porcelain is highly resistant to mechanical and electrical factors and heat. It is widely used to manufacture low- and high-voltage insulators.

6. Mica is a mineral of crystalline structure, which can easily be split into flakes (known as splittings). It is highly resistant to heat, moisture, and has excellent insulating properties (see Table 1-1).

Micanite (which is a trade name) is composed of mica splittings bonded together by varnish or resin. It is made into various washers and used in the manufacture of moulded products.

7. Paraffin wax is a substance obtained by the distillation of petroleum. It does not absorb water and melts at 55°C. Paraffin wax is used for impregnating paper, pressboard, wood, etc., so as to reduce their water absorption.

Chapter Two Direct-Current Circuits

2-1. Electric Current

In the absence of an external electric field, free electrons in a metal conductor move at random and the electric charge transferred through the conductor is on the average equal to zero.

In the presence of an electric field $\mathscr{E} = V/l$ directed along the wire, the free electrons are accelerated in the direction opposite to the field. As a result, the free electrons otherwise moving about at random are all set in a uniformly accelerated motion in the above direction. An electron remains in accelerated motion until it collides with an ion in the crystal lattice of the wire metal, after which it resumes its motion. Thus, the presence of a longitudinal electric field causes a certain electric charge to flow through any cross-section of a wire. The flow of charged particles in a conductor under the influence of an electric field is called an *electric current.*

Electrolytes (solutions of acids, salts and alkalis) are all *ion* (or *ionic*) conductors. Some molecules of the electrolyte are decomposed by the solvent into positive and negative ions which, like electrons in a metal, move about in the conductor. In ionic conductors, molecules of hydrogen and metals make up positive ions; molecules of nonmetallic radicals form negative ions.

Let us apply a voltage to two electrodes immersed into an electrolyte (Fig. 2-1). The electric field will cause positive ions to move to the negative electrode (cathode) and negative ions to the positive electrode (anode). This motion of positive and negative ions in the electrolyte caused by the electric field is an electric current. When they reach the

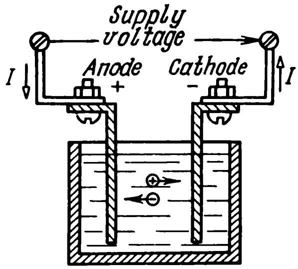


Fig. 2-1. Flow of current in an electrolyte

electrodes, the ions either deposit at the electrodes or react with them. At the anode, the negative ions of the electrolyte give up their electrons, and these then move around the circuit. At the cathode, the positive ions of the electrolyte combine with the free electrons coming from the circuit. Thus, in the wires leading from the power source, the free electrons move towards the cathode.

Current flow in electrolytes results from the transfer of the constituent particles of their substance.

According to Faraday's law, *the amount of any substance G deposited or dissolved at the electrodes is proportional to the quantity of electricity passing through the electrolyte*, that is:

$$G = cQ \tag{2-1}$$

Here, "*c*" ("*z*" in Britain) is the constant of proportionality, called the *electrochemical equivalent* of the substance, which is the weight of an element or group of elements set free by the passage of one coulomb of electricity. Every material has an electrochemical equivalent weight of its own; for example, 1.118 mg/coulomb for silver or 0.329 mg/coulomb for copper.

The decomposition of electrolytes due to the passage of a current is called *electrolysis*. It is widely used to obtain pure metals (such as copper or aluminium).

The rate of current flow is defined in terms of the strength of current.

Current strength, or simply *current*, *is equal to the quantity of electricity flowing through the cross-section of a conductor per unit time.* A current that does not change in either magnitude or direction is termed *direct current* (designated *I*).

If a charge Q flows in a conductor during a time t, current will be

$$I = Q/t \tag{2-2}$$

A current that changes with time is termed *alternating current* (designated i).

In the International System the unit of current is the ampere (A).

Current is equal to one ampere, if one coulomb flows through the cross-section of a conductor per second, that is:

$$1\ \text{A} = 1\ \text{coulomb}/1\ \text{s}$$

The positive direction of current flow is the direction for the motion of positive particles, or the direction opposite to that for the motion of electrons.

2-2. Electric Circuit and Its Components

An *electric circuit* is an assemblage of devices intended to produce and utilize electric current. It consists essentially of a power source, utilizing devices or loads, and wires which connect the source and loads.

Figure 2-2 shows an elementary circuit in schematic form. It is customary to divide an electric circuit into an internal and an external circuit. The internal circuit consists of energy source E, while the external circuit ($ABCD$) includes load r and wires AB and CD.

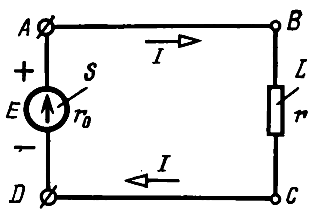


Fig. 2-2. Diagram of an electric circuit

Table 2-1 gives the symbols used in circuit diagrams in the USSR.

Table 2-1
Symbols Used in Circuit Diagrams

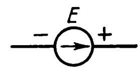
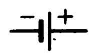
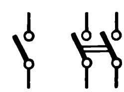
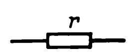
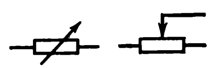
Circuit Component	Symbol
Electric energy source, e.m.f. source	
D-c generator	
D-c motor	
Chemical power source (primary or storage cell)	
Electric lamp	
Wire, cable, bus	
Electric connection, removable and permanent connection, removable connection	
Single- and double-pole switches	
Fuse	
Load, resistor	
Rheostat or adjustable resistor	

Table 2-1 (cont.)

Circuit Component	Symbol
Ammeter, voltmeter, wattmeter	Ⓐ Ⓥ Ⓦ

Power sources may be electric (rotary) *generators* converting mechanical into electric energy, *storage* or *primary cells* which convert chemical energy, etc.

Loads include, for example, *electric motors* which convert electric into mechanical energy, *electrolytic cells* intended to obtain pure metals (they convert electric into chemical energy), *incandescent lamps* and *heaters* which convert electric energy into light and heat, etc.

In a power source, some form of energy is converted into electric energy. Due to the work done by external (nonelectric) forces, each unit charge moving in a conductor gains some amount of energy. *The amount of energy gained by a unit charge in a conductor is called electromotive force* (abbreviated emf). When the external circuit is open-circuited, the emf (designated by the symbol E) is equal to the voltage across the terminals of the power source.

Utilizing devices or loads convert electric energy into thermal, mechanical or chemical energy. Now the *voltage V across the terminals of a load shows how much electric energy is spent per unit charge.*

The difference between the emf E and the voltage V is the energy that is converted into heat (that is, lost) in moving a unit charge in the power source. It is called the *voltage drop* (symbolized V_0). So

$$E - V = V_0, \text{ or } E = V + V_0 \tag{2-3}$$

Electric power is delivered from power sources to loads by wires. Within short stretch of wire, power loss may sometimes be ignored as has been done above.

Wires can be aluminium or copper, insulated or bare.

In addition to the three components which have been considered, electric circuits use various switches, protection

devices (fuses and relays), and instruments (such as ammeters, voltmeters, and wattmeters).

2-3. Ohm's Law

The ratio of a current I to the cross-sectional area S of a conductor is known as *current density* (designated by the letter δ):

$$\delta = I/S \tag{2-4}$$

Thus, the current density in a conductor is determined by the amount of charge passing through the unit cross-sectional area of the conductor per second, which is proportional to the rate of motion of charged particles along the wire. The rate of motion of the particles is proportional to the field forces, that is, the electric field strength. Thus, the current density* in a conductor is proportional to the electric field strength

$$\delta = \gamma \mathscr{E} \tag{2-5}$$

where $\gamma = \delta/\mathscr{E}$ is the constant of proportionality called *electric conductivity*.

Recalling $\mathscr{E} = V/l$, we may write

$$\delta = \gamma V/l$$

Multiplying the left- and right-hand sides of the last equation by the cross-sectional area of the conductor gives

$$I = V\gamma S/l = V/r \tag{2-6}$$

where the quantity

$$r = l/(\gamma S) \tag{2-7}$$

is called the *electric resistance of a conductor*.

From Eq. (2-6) it follows that *the current flowing in a wire is directly proportional to the applied voltage and inversely proportional to the wire resistance*.

Equation (2-6) is known as *Ohm's law*, one of the basic laws of electrical engineering, which is widely used in circuit calculations.

---

* In the International System, the unit of current density is 1 A/m^2 or 1 $A/mm^2 = 10^6$ A/m^2.

If a load has a resistance r (Fig. 2-2) and the associated power source has an internal resistance r_0, we may write for the circuit shown in Fig. 2-2

$$E = V + V_0 = Ir + Ir_0 = I\,(r + r_0) \qquad (2\text{-}8)$$

whence Ohm's law for a complete electric circuit is

$$I = \frac{E}{r + r_0} \qquad (2\text{-}9)$$

The voltage across the circuit (Fig. 2-2), with load connected, that is, under load, is given by

$$V = E - V_0 = E - Ir_0 \qquad (2\text{-}10)$$

The voltage across the same circuit with load disconnected, or at no load, when $I = 0$, is equal to the emf of the power source.

Example 2-1. Find the current in a filament lamp having a resistance of 440 ohms and connected to a 110-V supply line.

Solution.

From Ohm's law

$$I = V/r = 110/440 = 0.25 \text{ A}$$

Example 2-2. Find the voltage across a heater having a resistance of 44 ohms, if the current flowing in it is 5 A.

Solution.

The voltage across the heater is

$$V = Ir = 5 \times 44 = 220 \text{ V}$$

2-4. Electric Resistance and Conductance

When electric current flows in a conductor, that is when a directed movement of free electrons (or ions) takes place, the electrons (ions) run into atoms or molecules of the conductor, and these offer opposition to their flow, called the *resistance of the conductor*. In accordance with Eq. (2-6), the resistance is defined as the ratio between the voltage applied to and the current flowing in the conductor:

$$r = V/I \qquad (2\text{-}11)$$

In the International System, the unit of resistance is the ohm:

$$1 \text{ V}/1 \text{ A} = 1 \text{ V/A} = 1 \text{ ohm}$$

A conductor has a resistance of 1 ohm if a current of 1 A flows in it at a voltage of 1 V.

The reciprocal of conductivity is termed *resistivity* (designated ρ):

$$\rho = 1/\gamma \tag{2-12}$$

The resistivity of a conductor, like its conductivity, depends on the properties and temperature of its material.

Replacing the conductivity γ by the resistivity in Eq. (2-7), we obtain

$$r = l/(\gamma S) = \rho l/S \tag{2-13}$$

Hence

$$\rho = 1/\gamma = rS/l \tag{2-14}$$

The reciprocal of resistance is known as *conductance*, g

$$g = 1/r = \gamma S/l = I/V \tag{2-15}$$

The unit of conductance is the *siemens* (symbolized as S):

$$1 \text{ S} = 1/1 \text{ Ohm}$$

In the International System, the unit of resistivity is ohm m because

$$[\rho] = [rS/l] = \text{ohm} \times \text{m}^2/\text{m} = \text{ohm} \times \text{m}$$

With such a unit of measurement, the resistivities of metals will be expressed by inconveniently small numbers. Therefore, the length of a wire is taken in metres and its cross-sectional area in mm^2, which gives the following unit of resistivity

$$[\rho] = [rS/l] = \text{ohm} \times \text{mm}^2/\text{m}$$

or ohm $\times$ mm^2/m $= 10^{-6}$ ohm $\times$ m $= 1$ μohm $\times$ m.

In the International System, the unit of conductivity is

$$[\gamma] = \text{siemens/metre} = \text{S/m}$$

or, when taking the respective wire dimensions in metres and millimetres

$$[\gamma] = [1/\rho] = \text{m}/(\text{ohm} \times \text{mm}^2)$$

Table 2-2 summarizes the resistivity of some materials at a temperature of 20°C.

Table 2-2

Properties of Some Conductors Used in Electrical Engineering

Material	Melting point, °C	Resistivity at 20°C, μohm × m = = 10^{-6} ohm×m	Average value of temperature coefficient of resistance (from 0 to 100° C), °C^{-1}
Aluminium	657	0.029	0.004
Bronze	900	0.21-0.04	0.004
Tungsten	3370	0.055	0.00464
Constantan	1200	0.4-0.51	0.000005
Brass	900	0.07-0.08	0.002
Manganin	960	0.42	0.00006
Copper	1083	0.0175	0.004
Nichrome	1360	1.1	0.00015
Steel	1400	0.13-0.25	0.006
Fechral	1450	1.2	0.00005

In dielectrics (insulating materials), current may flow through the volume of the material, in which case we have *volume current* I_V, and on its surface, in which case we have *surface current* I_S. Due to this fact, two concepts are distinguished: *volume resistance* r_V and *surface resistance* r_S.

The volume current of a dielectric is

$$I_V = V/r_V = \frac{V}{\rho l/S}$$

The volume resistivity ρ_V of a dielectric is equal to the resistance of a specimen 1 m^2 in cross-section and 1 m long. Accordingly, the unit of volume resistivity is the same as in the case of metals: ohm m, because

$$[\rho] = [r_V S/l] = \text{ohm} \times \text{m}^2/\text{m} = \text{ohm} \times \text{m}$$

The volume resistivity of a cube with an edge of 1 cm is $[\rho]$ = 1 ohm × cm = 10^{-2} ohm × m.

Table 1-1 gives the volume resistivity ρ of some materials.

Surface current is written

$$I_S = \frac{V}{r_S} = \frac{V}{\rho_S l/d} \tag{2-16}$$

Fig. 2-3. Resistor

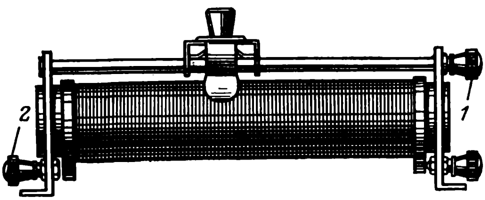


Fig. 2-4. Rheostat

The surface resistivity ρ_S of a dielectric is equal to the resistance of a specimen whose surface area is 1 m wide (d) and 1 m long (l). Thus, the unit of surface resistivity is the ohm, because

$$[\rho_S] = [r_S d/l] = \text{ohm} \times \text{m/m} = \text{ohm} \qquad (2\text{-}17)$$

Devices intended for connection into an electric circuit to limit or regulate the circuit current are called *resistors* (Fig. 2-3) or *rheostats* (Fig. 2-4). There are wire-wound and composition resistors, fixed or adjustable.

Heaters, rheostats and resistors use wire made of materials having high resistivity (see Table 2-2). With such materials it is possible to obtain the necessary resistance using a short length of wire which is wound on a ceramic or other insulating core (or former). A slider or contact arm *1* moving across the winding changes the resistance between terminals *1* and *2*.

In composition resistors, the resistance element is made as a rod or film applied to the surface of an insulating core (former).

2-5. Relationship between Resistance and Temperature

With increasing temperature, free electrons encounter atoms more frequently. Therefore, the average rate of the directed electron flow decreases, and this leads to an increase in the resistance of the conductor.

On the other hand, the number of free electrons and ions in the unit volume rises with temperature, so the resistance of the conductor drops.

Accordingly, the resistance of metals increases with rising temperature, whereas that of carbon and electrolytes decreases. There are several materials, however, such as metal alloys (manganin), whose resistance remains practically unaffected by the temperature rise.

At insignificant variations in temperature (from 0 to 100°C), the relative increment of resistance $\Delta r/r$ corresponding to 1°C rise in temperature, which is termed *temperature coefficient of resistance* (designated by the letter α), remains constant for most metals.

Designating the resistance at temperatures Θ_1 and Θ_2 by r_1 and r_2, respectively, we may write the relative increment of resistance at the temperature rise from Θ_1 to Θ_2 as

$$\frac{\Delta r}{r_1}=\frac{r_2-r_1}{r_1}=\alpha(\Theta_2-\Theta_1) \tag{2-18}$$

Hence

$$r_2=r_1+r_1\alpha(\Theta_2-\Theta_1)=r_1[1+\alpha(\Theta_2-\Theta_1)] \tag{2-19}$$

The temperature coefficient of resistance for some materials is given in Table 2-2.

From Eq. (2-18) it follows that

$$\Theta_2=\frac{r_2-r_1}{r_1\alpha}+\Theta_1 \tag{2-20}$$

Equation (2-20) makes it possible to determine the temperature of a wire (winding) Θ_2 by measuring its resistance r_2 with r, Θ_1 and α specified.

Example 2-3. Find the resistance of a copper overhead line 400 m long and 10 mm² $= 10\times10^{-6}$ m² in cross-section at temperatures of +20 and −10°C.

Solution.

The resistance of the line wires at +20°C is

$$r_1=\rho\frac{2l}{S}=0.0175\times10^{-6}\frac{2\times400}{10\times10^{-6}}=1.4 \text{ ohms}$$

The resistance at −10°C is

$$r_2=r_1[1+\alpha(\Theta_2-\Theta_1)]=1.4[1+0.004(-30)]$$
$$=1.568 \text{ ohms}$$

Example 2-4. The copper winding of an electric motor has a resistance (r_1) of 1.2 ohms at $\Theta_1 = +20°C$. After an hour of operation, the same winding showed a resistance (r_2) of 1.4 ohms. Find the temperature of the motor winding after one hour of operation.

Solution.

$$\Theta_2 = \frac{r_2 - r_1}{\alpha r_1} + \Theta_1 = \frac{1.4 - 1.2}{0.004 \times 1.2} + 20°C$$
$$= \text{(approx.)}\ 62°C$$

2-6. Conductor Materials

Conductor materials used in electrical engineering may be divided into two groups.

The first group embraces materials having low resistivity. They should have a low temperature coefficient of resistance, good mechanical strength and resistance to corrosion.

Copper is commonly used owing to its low resistivity (see Table 2-2), sufficient mechanical strength, good workability, and resistance to corrosion. It is made into wires for various purposes, buswork, or straps. Copper obtained by electrolysis contains not over 0.1% impurities.

Besides refined copper, electrical engineering uses alloys of copper, cadmium, beryllium and zinc known as bronzes and brasses.

Aluminium, although its electrical and mechanical properties are inferior to those of copper (see Table 2-2), has gained wide spread. When copper wires are replaced by aluminium ones having the same resistance and length, the cross-sectional area of the aluminium conductors is increased by 60%, but their mass is reduced by 52%.

Overhead lines use stranded steel-cored conductors which consist of an envelope of aluminium wires around a core of one or more high-tensile galvanized steel wires.

Steel, due to its high resistivity, is used only for low power and communication lines.

The second group includes materials with high resistivity, such as Nichrome which is an alloy of nickel, chromium and iron, and Fechral which is an alloy of iron, chromium and aluminium. Owing to their good heat resistance, they are used to manufacture heaters, rheostats, etc.

Manganin is an alloy of 86% copper, 12% manganese and 2% nickel. Due to its low temperature coefficient of resistance and high resistivity, manganin is used in instrumentation (as standard resistors, etc.).

Carbon products may be of two varieties, graphite and carbon. These products include brushes for electric machines, electrodes, and composition resistors.

2-7. Energy and Power

In a closed electric circuit (Fig. 2-2), the emf of the power source causes electric charges to move continuously.

By the definition of emf given in Sec. 2-2, the work done by the external forces in moving an electric charge Q in the power source, or the electric energy obtained by the conversion of some other energy, is

$$W_s = EQ = EIt \tag{2-21}$$

Since energy is the ability to do work, the above expression also gives energy.

By the law of conservation of energy, the electric energy generated by a source during a time t is converted into other types of energy in the external circuit during the same time.

The voltage across the terminals of a source being $V_{AD} = V_{BC} = V$, the electric energy spent in the external source is

$$W = VQ = VIt \tag{2-22}$$

Some energy is spent (lost) in the power source as heat:

$$W_0 = W_s - W = (E - V)\, It = V_0 It$$

As has been given in Sec. 2-2, the difference between the emf of the source E and the voltage across its terminals V gives the internal voltage drop:

$$V_0 = E - V \tag{2-23}$$

The ratio of work W to the time t during which the work is done is termed *power* and designated by the letter P:

$$P = W/t \tag{2-24}$$

Thus, *power is the time rate of doing work or the time rate of transferring or transforming energy.*

The rate at which mechanical or other energy is converted into electric energy in a generator (power source) is called *source power*:

$$P_s = EIt/t = EI \tag{2-25}$$

The rate at which electric energy is converted into other forms in the external circuit is termed *load power*

$$P = VIt/t = VI \tag{2-26}$$

The power associated with the loss of energy in a generator is known as *power loss*:

$$P_0 = W_0/t = V_0It/t = V_0I \tag{2-27}$$

By the law of conservation of energy, *the power of a generator is equal to the sum of load powers it sustains and its power loss*

$$P_s = P + P_0 \tag{2-28}$$

In the International System, the unit of power, called the watt (W), is equal to one joule per second or to the power at which 1 joule of electric energy is converted into some other energy during one second, that is

$$1 \text{ W} = 1 \text{ J}/1 \text{ s}$$

or

$$1 \text{ J} = 1 \text{ W} \times 1 \text{ s} = 1 \text{ W s}$$

From Eq. (2-26) it follows that

$$[P] = [VI] = \text{VA}$$

so,

$$1 \text{ W} = 1 \text{ V} \times 1 \text{ A}$$

that is, the watt is the power at a current of 1 A and a voltage of 1 V.

Example 2-5. Find the current of an electric motor consuming 10 kW and connected in a 225-V supply line.

Solution.

Power is defined as

$$P = VI$$

Hence the current is

$$I = P/V = 10{,}000/225 = 44.4 \text{ A}$$

Example 2-6. The current in a heater connected in a 220-V supply line is 5 A. Find the power of the heater and the cost of the energy spent during 4 hours of operation. The cost of 1 kWh is 4 kopecks.

Solution.

The power of the heater is

$$P = VI = 220 \times 5 = 1100 \text{ W} = 1.1 \text{ kW}$$

The electric energy spent by the heater is

$$W = Pt = 1100 \times 4 = 4400 \text{ Wh} = 4.4 \text{ kWh}$$

and its cost is

$$4 \times 4.4 = 17.6 \text{ kopecks}$$

2-8. Conversion of Electric Energy into Heat

When electric current flows in a conductor having resistance r, charged particles run into ions and molecules of the conductor material. In so doing, the moving particles transfer their energy to the ions and molecules, due to which fact the conductor heats.

The rate at which electric energy is converted into thermal energy is given by

$$P = VI$$

Recalling that $V = Ir$, we obtain

$$P = I^2r, \quad \text{or} \quad P = V^2/r \qquad (2\text{-}29)$$

The amount of electric energy converted into thermal energy during a time t is

$$W = Pt = I^2rt$$

Since the unit of energy and unit of heat in the International System is the joule, the heat generated in the resistance r is

$$Q = I^2rt \qquad (2\text{-}30)$$

This relationship was experimentally discovered simultaneously by H.F.E. Lenz, a Russian academician, and Joule,

H. F. E. Lenz (1804-1865)

an English scientist, in 1844. It is called the Joule-Lenz* law and reads as follows:

The amount of heat produced in a conductor by the passage of electric current is proportional to the squared current, the resistance of the conductor and the time of current flow.

The conversion of electric energy into heat is utilized by electric ovens and various heaters. In electric machines and apparatus, however, this conversion leads to losses of energy and reduces their efficiency. In these units, heat build-up limits their load capacity and causes overloads. An overload is usually accompanied by a temperature rise which may damage the insulation or cut down the service life of the equipment.

Example 2-7. Find the amount of heat produced in a device during 1 hour of operation. The resistance of the device is 88 ohms and the voltage across its terminals is 220 V.

Solution.

The current in the device is

$$I = V/r = 220/88 = 2.5 \text{ A}$$

The amount of heat produced in the device is

$$Q = I^2rt = 2.5^2 \times 88 \times 1 \times 3600 = 1{,}980{,}000 \text{ J} = 1.98 \text{ MJ}$$

2-9. Electric Load on Wires and Overload Protection

Let a wire be heated by a current flowing in it. At first, the temperature of the wire is equal to the ambient temperature. So the heat produced by the current flow is spent to heat the wire. As the temperature of the wire increases, more heat is given up by the wire to the atmosphere, because the

* In the US and UK literature of the subject this law is known as Joule's law.— *Translator's note.*

difference in temperature between the wire and the surroundings increases. As a result, the rise in the temperature of the wire slows down. Finally, the wire temperature reaches what is known as a steady-state value, when the heat produced by the current flow is equal to the heat given up to the atmosphere. The time necessary to reach this value varies with different devices: the filament of a lamp reaches this temperature in a split second; an electric machine does so in several hours.

It is safe to let wires be heated to certain temperatures (65 to 80°C), determined by the properties of the insulation or wires. The current at which the maximum safe (allowable) temperature is reached is called the *maximum safe* (allowable) *current of a wire* (Table 2-3).

Table 2-3
Maximum Safe Current (Load) of Insulated Wires

Cross-sectional area of a wire, mm^2	0.5	0.75	1.0	1.5	2.5	4	6	10	16	25	35	50
Max. safe current, A	$\frac{11}{—}$	$\frac{15}{—}$	$\frac{17}{—}$	$\frac{23}{—}$	$\frac{30}{24}$	$\frac{41}{32}$	$\frac{50}{39}$	$\frac{80}{55}$	$\frac{100}{80}$	$\frac{140}{105}$	$\frac{170}{130}$	$\frac{215}{165}$

Note. The numerators give loads for copper wires, the denominators for aluminium wires.

This table can be used to choose the cross-sectional area of wires on the basis of maximum safe temperature. The table gives the maximum safe (operating) current, I_s, for standard insulated wires. A wire must be chosen to have such a cross-sectional area that its maximum safe continuous current is equal or a little greater than the maximum design (or rated) current:

$$I_s \geqslant \bar{I}_d$$

The connection of two wires at different potentials through a low-value resistance is called a *short-circuit*.

The short-circuit current is tens of times the nominal current of a device, and it may cause mechanical or thermal damage to its parts.

The components of an electric circuit are protected against overcurrents and short-circuits by fuses (Sec. 11-10) or relays (Sec. 11-13). A fuse consists of a piece of wire which melts at a certain overcurrent and breaks the circuit.

2-10. Voltage Loss in Wires

When electric energy is transferred over short wires, their resistance may be ignored. In the case of long wires (longer than 10 m), their resistance cannot be neglected, because it causes a noticeable voltage drop:

$$\Delta V = Ir = I\frac{2l}{\gamma S} \tag{2-31}$$

In this case, the term "voltage drop" refers to the difference between the voltages at the ends of a line (Fig. 2-5), $V_1 - V_2$, given by

$$V_1 - V_2 = \Delta V = Ir \tag{2-32}$$

and usually referred to as voltage loss.

With a constant voltage at the start of the line, the voltage at the finish of the line, that is, at load, varies from $V_2' = V_1$ at $I = 0$ to $V_2'' = V_1 - \Delta V$ under load.

Variations in voltage should not exceed —2.5 to +5 per cent of the nominal value for lighting accessories and ±5 per cent or sometimes ±10 per cent for power equipment. Voltage loss in a line should not exceed the same values.

When the allowed voltage loss in the line is specified in advance, we may determine the necessary cross-sectional area of the line wires, using Eq. (2-31):

$$S = 2Il/\gamma\Delta V \tag{2-33}$$

The cross-sectional area found by Eq. (2-33) and rounded to the nearest standard value should be checked for maximum safe current (see Table 2-3).

The power loss in a line is given by the product of voltage loss and current, that is

$$\Delta P = I\Delta V = I^2r$$

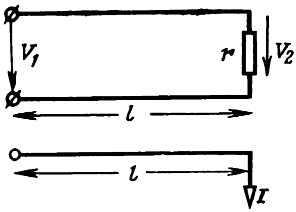


Fig. 2-5. Two-wire line loaded at the far end

The efficiency of a line

$$\eta = \frac{P_2}{P_1} = \frac{P_1 - \Delta P}{P_1} = 1 - \frac{\Delta V}{V_1}$$

decreases with rising load.

At voltage losses of 2 to 5 per cent, the efficiency of a line is 98-95 per cent.

2-11. Kirchhoff's First (Current) Law

If three or more wires (branches) (Fig. 2-6) meet at one point in a circuit, this point is called a *circuit junction* or a (circuit) *node*.

When there is a constant current in a circuit, none of its nodes can accumulate electric charges because it would change the potentials at those nodes. Therefore, the electric charges entering a junction (node) per unit time are equal to the charges leaving this junction (node) during the same period. This relationship, known as *Kirchhoff's first* (current) *law*, is stated as follows: *the sum of the currents entering a circuit node is equal to the sum of the currents leaving it.*

For example, for node A we may write:

$$I_1 + I_2 = I_3 + I_4 + I_5$$

or

$$I_1 + I_2 + (-I_3) + (-I_4) + (-I_5) = 0$$

so the final expression is

$$\Sigma I = 0 \tag{2-34}$$

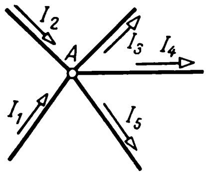


Fig. 2-6. Node (junction point) of an electric network

or, in words, *the algebraic sum of the currents in a circuit node is zero.* The currents leaving a node are considered negative.

2-12. Series Combination of Resistances (Loads)

In a series combination, the finish of one resistance (or load) is connected to the start of another, and so on (Fig. 2-7). When connected to a power source, they conduct the same current.

The energy spent to move a unit charge around the complete circuit is equal to the sum of the energies spent to move the same charge in all elements of the circuit, or, in other words, *the voltage across the circuit is equal to the sum of the voltages across the individual resistances*:

$$V = V_1 + V_2 + V_3 \tag{2-35}$$

Dividing the left- and right-hand sides of the equation by the circuit current gives

$$V/I = V_1/I + V_2/I + V_3/I$$

Hence

$$r = r_1 + r_2 + r_3 \tag{2-36}$$

The resistance r is called *the equivalent resistance of the circuit.* When the resistances of the circuit are replaced by its equivalent resistance at the same voltage, the circuit current remains unchanged.

Thus, *the equivalent resistance of a series combination of loads* (for example, resistors) *is equal to the sum of the resistances of the individual loads* (resistors).

The voltages across the resistors are determined by the following expressions: $V_1 = Ir_1$; $V_2 = Ir_2$; $V_3 = Ir_3$. Therefore, $V_1/V_2/V_3 = r_1/r_2/r_3$, or *the voltages across resistors con-*

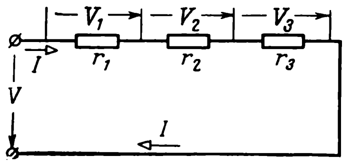


Fig. 2-7. Resistors connected in series

nected in series are proportional to their resistances. The circuit consisting of wires and the load considered in Sec. 2-10 (Fig. 2-5) is also a series combination of resistances.

Example 2-8. The winding of an electric motor with a resistance (r_1) of 24 ohms and a rheostat with resistance r_2 which can be adjusted from 0 to 96 ohms are connected in series to a 120-V supply line. Find the limits of the adjustment range for the circuit current.

Solution.

The equivalent resistance of the circuit is: $r = r_1 + r_2$. At $r_2' = 0$, the circuit current will be

$$I' = V/(r_1 + r_2') = 120/(24 + 0) = 5 \text{ A}$$

At $r_2'' = 96$ ohms, the circuit current will be

$$I'' = V/(r_1 + r_2'') = 120/(24 + 96) = 1 \text{ A}$$

2-13. Parallel Combination of Resistances (Loads)

In a parallel combination, resistances (loads) meet at common junction points, or terminals, at both ends, as shown in Fig. 2-8. Thus, there may be several parallel branches between the two circuit terminals.

The voltages across the resistances (loads) are the same and equal to the voltage between the terminals

$$V = V_1 = V_2 = V_3 \tag{2-37}$$

The currents in the resistances (loads) may be derived from Ohm's law:

$$I_1 = V_1/r_1 = Vg_1; \quad I_2 = V_2/r_2 = Vg_2; \quad I_3 = V_3/r_3 = Vg_3 \tag{2-38}$$

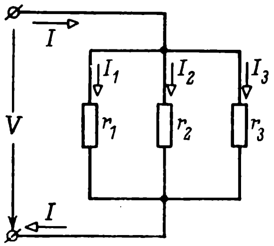


Fig. 2-8. Resistors connected in parallel

Hence,

$$I_1/I_2/I_3 = \frac{1}{r_1}\Big/\frac{1}{r_2}\Big/\frac{1}{r_3} = g_1/g_2/g_3 \qquad (2\text{-}39)$$

that is, *the currents in parallel branches are inversely proportional to their resistances or directly proportional to their conductances.*

The resistances of several individual loads (for example, resistors) may be replaced by the equivalent resistance r whose current I should be equal to the sum of branch currents ΣI, at the same terminal voltage. Thus, the equivalent resistance is

$$r = V/I = V/(I_1 + I_2 + I_3)$$

and the equivalent conductance is

$$\begin{aligned} 1/r &= (I_1 + I_2 + I_3)/V = 1/r_1 + 1/r_2 + 1/r_3 \\ &= g_1 + g_2 + g_3 = g \qquad (2\text{-}40) \end{aligned}$$

This equation states that the equivalent conductance of a parallel combination of resistances is equal to the sum of the conductances of the individual resistances.

From Eq. (2-40), we can obtain an expression for the equivalent resistance of a parallel combination. For example, for three resistances:

$$1/r = 1/r_1 + 1/r_2 + 1/r_3 = (r_1r_2 + r_1r_3 + r_2r_3)/r_1r_2r_3$$

Hence

$$r = r_1r_2r_3/(r_1r_2 + r_1r_3 + r_2r_3) \qquad (2\text{-}41)$$

When the branch resistances are equal,

$$r = r_1^3/3r_1^2 = r_1/3$$

If there are n branches and the resistances in parallel are of the same value r_1, the equivalent resistance of the parallel circuit will be

$$r = r_1/n \tag{2-42}$$

In the case of two resistances (branches) connected in parallel, the equivalent resistance of the parallel circuit will be, in accordance with Eq. (2-40), as follows:

$$r = r_1 r_2/(r_1 + r_2) \tag{2-43}$$

Loads (electric motors, incandescent lamps, heaters) are designed for operation at a constant nominal voltage; therefore, they are connected in parallel.

Example 2-9. An electric motor consuming a power of 5.5 kW and eleven 100-W incandescent lamps are connected to a 220-V supply line. Find the current in the wires.
Solution.
The current of the motor is

$$I_1 = P_1/V = 5500/220 = 25 \text{ A}$$

The current drawn by the lamps is

$$I_2 = P_2/V = 100 \times 11/220 = 5 \text{ A}$$

The current in the wires is

$$I = I_1 + I_2 = 25 + 5 = 30 \text{ A}$$

Example 2-10. Find the equivalent resistance of ten 200-W incandescent lamps connected in parallel at a nominal voltage of 220 V.
Solution.
The resistance of one lamp is

$$r_l = V_l^2/P_l = 220^2/200 = 242 \text{ ohms}$$

The equivalent resistance of ten lamps is

$$r = r_l/n = 242/10 = 24.2 \text{ ohms}$$

2-14. Series-Parallel Combination of Resistances (Loads)

In a series-parallel combination, some of the resistances (or branches) are in series and some in parallel.

A typical example of a series-parallel combination is furnished by several parallel resistances connected in series with the supply line resistance (Fig. 2-9). For a specified terminal voltage and known branch resistances, we "solve" this circuit (such as in analysis or synthesis) by finding the currents and voltages of all parts or sections of the circuit. The circuit shown in Fig. 2-9 consists of two sections connected in series: BC consisting of three parallel branches, and AB having a resistance r_1.

The conductance of multi-branch (parallel) section BC is

$$g_{BC} = 1/r_2 + 1/r_3 + 1/r_4$$

so its resistance is

$$r_{BC} = 1/g_{BC}$$

The equivalent resistance of the complete circuit is

$$r = r_{AB} + r_{BC} = r_1 + r_{BC}$$

From Ohm's law, the circuit current is

$$I = V/r = V/(r_1 + r_{BC})$$

The voltages across sections AB and BC are:

$$V_{AB} = V_1 = Ir_1 \quad \text{and} \quad V_{BC} = Ir_{BC}$$

The currents in the parallel branches are

$$I_2 = V_{BC}/r_2; \quad I_3 = V_{BC}/r_3; \quad I_4 = V_{BC}/r_4$$

Example 2-11. Find the currents and voltages of all sections of the circuit shown in Fig. 2-9, if $V = 240$ V; $r_1 = 2.12$ ohms; $r_2 = 20$ ohms; $r_3 = 10$ ohms; $r_4 = 50$ ohms.

Solution.

The conductance of multi-branch (parallel) section BC is

$$g_{BC} = 1/r_{BC} = 1/r_2 + 1/r_3 + 1/r_4$$
$$= 1/20 + 1/10 + 1/50 = 0.17 \text{ S}$$

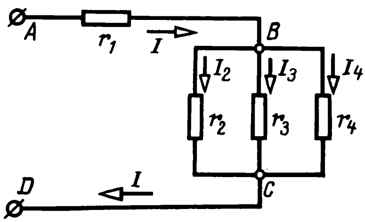


Fig. 2-9. Resistors connected in series-parallel

The resistance of the parallel section is

$$r_{BC} = 1/g_{BC} = 1/0.17 = 5.88 \text{ ohms}$$

The equivalent resistance of the complete circuit is

$$r = r_1 + r_{BC} = 2.12 + 5.88 = 8 \text{ ohms}$$

The circuit current is

$$I = V/r = 240/8 = 30 \text{ A}$$

The voltage across the first resistance is

$$V_1 = Ir_1 = 30 \times 2.12 = 63.6 \text{ V}$$

The voltage across the parallel section is

$$V_{BC} = Ir_{BC} = 30 \times 5.88 = 176.4 \text{ V}$$

The currents in the parallel branches are

$$I_2 = V_{BC}/r_2 = 176.4/20 = 8.82 \text{ A}, \quad I_3 = V_{BC}/r_3$$
$$= 176.4/10 = 17.64 \text{ A}$$
$$I_4 = V_{BC}/r_4 = 176.4/50 = 3.53 \text{ A}$$

2-15. Two Modes of Operation of Power Source

Figure 2-10 shows a circuit with two power sources having resistances r_{01} and r_{02}, respectively. The current in this circuit may be found by the *superposition theorem**, as the

* When a number of voltages are applied to the network simultaneously, the current that flows is the sum of the component currents that would flow if the same voltages acted individually.— *Translator's note.*

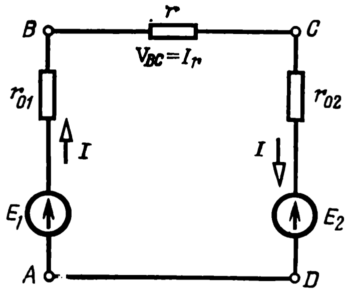


Fig. 2-10. Circuit containing two sources

algebraic sum of the currents generated by the individual sources. To begin with, assume that the emf of the first power source is $E_1 \neq 0$ and that of the second is $E_2 = 0$; then, conversely, we assume that $E_1 = 0$ and $E_2 \neq 0$.

In the first case, the circuit current which coincides in direction with emf E_1, can be expressed as:

$$I_1 = E_1/(r_{01} + r_{02} + r)$$

In the second case, the current flowing in the direction of emf E_2 is

$$I_2 = E_2/(r_{01} + r_{02} + r)$$

When the emfs exist simultaneously, that is, when $E_1 \neq 0$ and $E_2 \neq 0$, the circuit current can be determined by adding together the currents I_1 and I_2 (according to the superposition theorem):

$$I = I_1 + I_2 = (E_1 + E_2)/(r_{01} + r_{02} + r) \qquad (2\text{-}44)$$

If emfs E_1 and E_2 are aiding each other, the currents I_1 and I_2 flow in the same direction, too.

If E_1 and E_2 are opposing each other (as shown in Fig. 2-10), the circuit current is equal to the difference between the currents

$$I = I_1 - I_2 = (E_1 - E_2)/(r_{01} + r_{02} + r) \qquad (2\text{-}45)$$

that is, a current will only flow when $E_1 \neq E_2$ and its direction coincides with the direction of the higher emf. Assume that $E_1 > E_2$, then the direction of current I agrees with the direction of E_1 and is opposite to that of E_2. An electromotive force opposing the current is called the *counter-emf*.

In the resistance r (section BC, Fig. 2-10), electric energy is converted into thermal energy. The associated power is

$$P_{BC} = I^2 r$$

and the voltage drop is

$$V_{BC} = P_{BC}/I = Ir$$

In addition to thermal power $I^2 r_{02}$, there is also power $E_2 I$ which is generated in section CD because the electric forces do work in overcoming the counter-emf. This power is converted into chemical or mechanical power, depending on the type of power source used (a storage battery or an electric machine). Thus, the source with counter-emf E_2 dissipates electric energy, that is, acts as a load.

The power developed in section CD is

$$P_{CD} = E_2 I + I^2 r_{02}$$

and the voltage across this section is

$$V_{CD} = P_{CD}/I = E_2 + I r_{02} \qquad (2\text{-}46a)$$

Therefore, *the voltage across a power source acting as a load is equal to the sum of its emf and its internal voltage drop.*

In section BA, the emf E_1 coincides in direction with the current I, so the source (a storage battery or an electric machine) operates as a generator. Therefore, its emf is equal to the sum of the terminal voltage and the internal voltage drop (2-8):

$$E_1 = V_{BA} + V_0 = V_{BA} + I r_{01}$$

so the terminal voltage is

$$V_{BA} = E_1 - I r_{01} \qquad (2\text{-}46b)$$

This suggests that *the terminal voltage of a power source operating as a generator is equal to the difference between its emf and its internal voltage drop.*

According to Eq. (2-25), the power developed by a power source operating as a generator is

$$E_1 I = V_{BA} I + V_0 I = V_{BA} I + I^2 r_{01}$$

As we have seen, a power source may operate either as a generator or as a load. In the former case, its terminal voltage is less than its emf ($V < E$), and its current has the

same direction as its emf. In the latter case, its voltage is greater than the emf ($V > E$), and its current opposes its emf.

2-16. Kirchhoff's Second (Voltage) Law

A branch of an electric circuit or network is a section of the network between two adjacent junction points or nodes. A set of branches forming a closed path in a network is termed a *loop* or *mesh.*

Generally, an electric circuit, or network, may include several power sources and resistances connected arbitrarily, for example, as shown in Fig. 2-11.

Assume that both power sources operate as generators, that is, the currents agree in direction with the emfs. For them, the voltages across CA or, which is the same, across FG, are the same and given by Eq. (2-46):

$$V_{CA} = E_1 - I_1 r_1; \; V_{CA} = E_2 - I_2 r_2$$

where r_1 and r_2 include the internal resistances of the sources, r_{01} and r_{02}.

From a comparison of these two equations, we find that the following equation applies to the mesh $ABCDA$

$$E_1 - I_1 r_1 = E_2 - I_2 r_2$$

whence

$$E_1 - E_2 = I_1 r_1 - I_2 r_2$$

or generally

$$\Sigma E = \Sigma (Ir) \tag{2-47}$$

The above expression is Kirchhoff's second (voltage) law which reads as follows: *The algebraic sum of the emfs round a mesh in a network is equal to the algebraic sum of all the voltage drops around it.*

When writing equations according to this law, the emfs E whose directions coincide with the arbitrary direction of summation round a mesh are assigned the "+" sign, and the emfs with opposite directions are assigned the "—" sign.

The voltage drops Ir are taken with the "+" sign if the direction of summation round a mesh agrees with the dire-

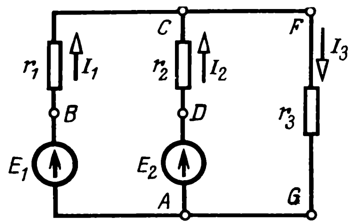


Fig. 2-11. Multi-mesh (multi-loop) circuit with two sources

ction of the current I in the resistance r. Otherwise, the voltage drops have the "—" sign.

When carrying out summation round the mesh $ABCDA$ clockwise, we may write (see Fig. 2-10)

$$E_1 + (-E_2) = Ir_{01} + Ir + Ir_{02}$$

which corresponds to what has been said in Sec. 2-15.

2-17. Solution of Complex Circuits

A *complex circuit*, or network, is one containing several loops or meshes with power sources and loads arranged in any arbitrary way, which cannot be reduced to a combination of series and parallel connections.

In addition to Ohm's law, the basic laws for circuit solution (analysis and synthesis) are two Kirchhoff's laws. Using them, one can find currents and voltages for all sections of any complex network.

One of the methods of solving complex networks is described in Sec. 2-15; this is the superposition theorem. This method is based on the principle that the current in any branch is the algebraic sum of the currents generated in it by each individual source of emf in the circuit.

Let us solve a complex circuit, using the node equations and the mesh equations or the equations of Kirchhoff's laws.

In order to find currents in all branches of a circuit, it is necessary to know the resistances of the branches and also the values and directions of all emfs.

Before writing the equations by Kirchhoff's laws, we should choose the directions of the branch currents and label them with arrows on the circuit diagram. If the chosen di-

rection of the current in any branch is opposite to its real direction, this current will appear negative after the equations have been solved.

The number of equations is equal to the number of unknown currents. The number of equations written by Kirchhoff's first (current) law should be equal to the number of nodes less one. The remaining equations are written by Kirchhoff's second (voltage) law. When applying this law it is advisory to choose the simplest networks, and each of them should contain at least one branch not used in the previous equations.

Here is an example of a complex circuit solved by Kirchhoff's two equations.

Example 2-12. Find the currents in each branch of the circuit shown in Fig. 2-11, if $E_1 = 246$ V, $E_2 = 230$ V, $r_1 = 0.3$ ohm, $r_2 = 1$ ohm, and $r_3 = 24$ ohms. The internal resistances of the sources may be ignored.

The arbitrarily chosen directions of the branch currents are shown in Fig. 2-11.

Solution.

There are three unknown currents, so three equations need be written.

For two nodes, one node equation is necessary. Let us write it for node C

$$I_1 + I_2 - I_3 = 0 \qquad (2\text{-}48)$$

To write the second equation, we carry out summation round the network $ABCFGA$ clockwise:

$$E_1 = I_1 r_1 + I_3 r_3 \qquad (2\text{-}49)$$

For the third equation, we carry out summation round the network $ADCFGA$ and write

$$E_2 = I_2 r_2 + I_3 r_3 \qquad (2\text{-}50)$$

Substituting the numerical values for the literals in Eqs. (2-49) and (2-50), we obtain

$$246 = 0.3I_1 + 24I_3 \qquad (2\text{-}51)$$

$$230 = 1I_2 + 24I_3 \qquad (2\text{-}52)$$

Substituting its expression from Eq. (2-48) for the current I_2 in the last equation, we obtain

$$230 = 1I_3 - 1I_1 + 24I_3 = -I_1 + 25I_3 \quad (2\text{-}52a)$$

Multiplying Eq. (2-52*a*) by 0.3 and adding to Eq. (2-51) we shall have

$$\begin{array}{rrr} & 69 = & -0.3I_1 + 7.5I_3 \\ + & 246 = & 0.3I_1 + 24I_3 \\ \hline & 315 = & 31.5I_3 \end{array} \quad (2\text{-}53)$$

whence the current in the third branch is

$$I_3 = 31.5/3.15 = 10 \text{ A}$$

The voltage across the third branch is

$$V_{FG} = I_3 r_3 = 10 \times 24 = 240 \text{ V}$$

The currents in the first and second branches are

$$I_1 = (E_1 - V_{FG})/r_1 = (246 - 240)/0.3 = 20 \text{ A}$$
$$I_2 = (E_2 - V_{FG})/r_2 = (230 - 240)/1 = -10 \text{ A}$$

The negative value of the current I_2 indicates that the actual direction of this current is opposite to that labelled in the diagram (Fig. 2-11). Thus, the power source E_1 generates energy and the power source E_2 dissipates it.

2-18. Chemical Sources of Current

(a) Primary Cells

There is always a certain difference in potential between an electrode and the electrolyte in which it is immersed, the magnitude of the difference being dependent on the material of the electrode and the composition of the electrolyte.

The *electrode potential* arises owing to the fact that chemical forces cause the electrode material to dissolve in the electrolyte (for example, zinc dissolves in a solution of sulphuric acid) and its positive ions pass into the electrolyte. More specifically, the prevalence of negative charges on the electrode and positive ones in the boundary layer of electrolyte adjacent to the electrode produces a double electric layer and, as a consequence, an electric field at the

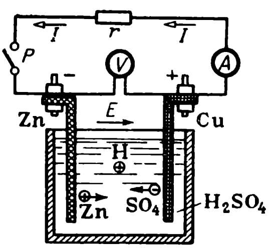


Fig. 2-12. Volta's cell and its connection in a circuit

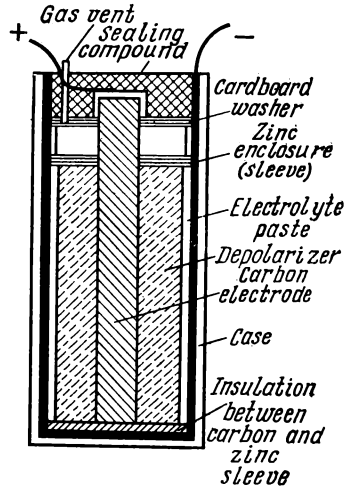


Fig. 2-13. Zinc-manganese cell of the cylinder type

electrolyte boundary, directed away from the electrolyte towards the electrode. This field prevents the positive ions of the electrode from passing into solution and balances the chemical forces that are responsible for the dissolution of the electrode. This gives rise to the electrode potential.

Placing two electrodes made of different metals into an electrolyte, we obtain a potential difference between them, or an emf, given by $E = \varphi_1 - \varphi_2$.

Thus, a device consisting of two dissimilar electrodes immersed in an electrolyte is a source of an emf, that is, a *primary current source or cell in which chemical energy irreversibly converts into electric energy.*

There is a great variety of primary cells. As an example, let us consider Volta's primary cell (Fig. 2-12). It uses a zinc plate (Zn) and a copper plate (Cu) as electrodes immersed in diluted sulphuric acid (H_2SO_4). The zinc electrode is charged negatively (it is called the cathode) and the copper electrode is charged positively (it is called the anode). The emf of this primary cell is about 1.1 V.

When the cell is loaded, that is, when an electric current flows in it, negative SO_4 ions and positive Zn ions approach

one another and combine to form molecules of zinc sulphate ($ZnSO_4$). At the same time, the positive ions of hydrogen attract electrons from the anode and form neutral atoms of hydrogen which, covering the surface of the anode, cause a rise in the internal resistance of the cell and reduce its emf. This is called *polarization*. The hydrogen layer over the anode can be removed by using *depolarizers*, the materials easily giving up oxygen (for example, manganese dioxide) which combines with hydrogen to form water.

The most widely used primary cell of nonpolarizing type is the Leclanché cell. The Leclanché cell is available in wet and dry forms. It can also be of a cylinder or a layer type.

The cylinder type cell is put into a zinc jar (Fig. 2-13) which also serves as the cathode; a carbon rod placed centrally in the jar serves as the anode. The depolarizer is a mixture of manganese dioxide, graphite and carbon black. The zinc jar is filled with an aqueous solution of ammonium chloride (popularly called sal ammoniac) pasted by starch. The emf of this cell is 1.5 V.

The long-time current allowed in the operation of a cell is called its nominal discharge current. The quantity of electricity that a cell is able to give up during discharge is called its capacity and is measured in ampere-hours (Ah). Dry cell batteries find wide use in radio engineering, wire communication equipment, flash lights, hearing aids, portable meters, etc.

(b) Storage Cells (Secondary Chemical Sources of Current)

The chemical sources of current which can be recharged after discharge, in which case electrical energy is converted into chemical energy, are called *storage* or *secondary cells*.

There are two distinct groups of storage cells: acid and alkaline. The acid group includes lead-type storage cells, while the alkaline group consists of nickel-cadmium, nickel-iron and silver-zinc type storage cells.

A *lead-acid storage cell* consists of two plate groups (Fig. 2-14) immersed in an electrolyte of 25-35% aqueous solution of sulphuric acid.

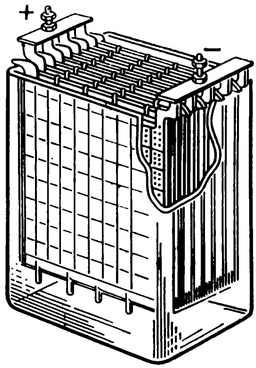


Fig. 2-14. Lead-acid storage battery

The positive plates are made of pure lead, and a highly developed surface is obtained by cutting, spinning or rolling upstanding fins or leaves on the faces of a pure lead blank. They can also consist of cast grids of lead filled with active material (lead peroxide).

The negative plates are formed by lead grids into which the active material (sponge lead) is pasted. After manufacture, the plates are given a treatment, called forming.

During discharge, that is, when a charged cell is connected to an external circuit, the discharge current flows and the cell operates as a power source. In this process, the active material of the positive plates (lead peroxide, PbO_2) and that of the negative plates (sponge lead, Pb) become lead sulphate, $PbSO_4$, and water. This reduces the concentration and conductivity of the electrolyte and the emf of the storage cell. The voltage (emf) of a cell rapidly falls from 2.2 V to 2 V, then slowly to 1.8 V. Lead storage cells should never be discharged below this value (called the *final voltage*), otherwise sulphatization of the plates (the formation of insoluble lead sulphate) may take place.

On charging, the charging current flows through a cell in the direction opposite to the discharge current, for which purpose the terminals of a battery charger are connected to the similar terminals of the cell.

During the charging, the reverse chemical reaction takes place, and lead peroxide and sponge lead are reduced on the electrodes. At first, the voltage rapidly rises to 2.2 V,

then more slowly to 2.3 V and, finally, to 2.6-2.7 V. At this voltage, intensive bubbling begins at the plates, as if the electrolyte were boiling, and the charging should be stopped.

The internal resistance of lead-acid cells is low, so short-circuit currents are intolerably high.

The capacity of a storage cell, as is that of a primary cell, is measured in ampere-hours during the time of normal discharge.

The ratio of the quantity given up by a storage cell during discharge to the quantity of electricity obtained on charging is called the *capacity* (ampere-hour) *efficiency* of the storage cell

$$\eta = Q_d/Q_{ch} \tag{2-54}$$

The capacity efficiency of lead storage cells is between 0.9 and 0.95.

The ratio of the energy obtained from a storage cell on discharge, W_d, to the energy spent to charge the cell, W_{ch}, is termed the *overall* (watt-hour) *efficiency* of the storage cell

$$\eta = W_d/W_{ch} \tag{2-55}$$

and its common value for lead-acid cells is anywhere between 0.75 and 0.8.

In order to prevent a storage cell from sulphatization, it is necessary to keep it charged, periodically check the level and density of the electrolyte, voltage under load, and recharge it when necessary.

Alkaline storage cells are called so because they use an electrolyte of 21% aqueous solution of potassium hydrate (KOH) or sodium hydrate (NaOH). They consist of two plates placed in a steel container holding an electrolyte (Fig. 2-15). The plates are formed by steel pockets filled with an active material and assembled in steel frames. The active material of nickel-cadmium storage cells is sponge cadmium and that of nickel-iron storage cells is sponge iron. Both types use the same active material of nickel hydrate $Ni(OH)_3$ on positive plates.

During discharge, the nickel hydrate changes to nickel hydroprotoxide, and the sponge cadmium (or iron) changes to cadmium hydroprotoxide (or ferrous hydroxide). On charging, the chemical reactions are reversed; therefore, the

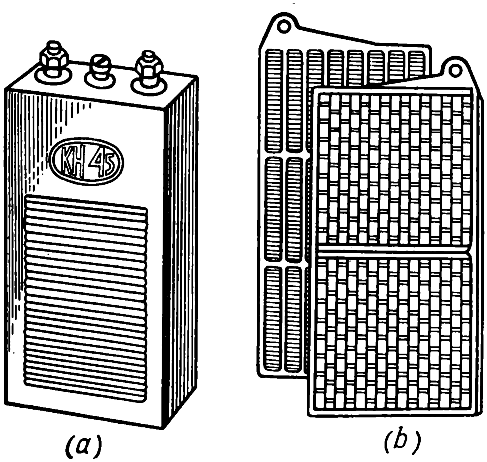


Fig. 2-15. Alkaline storage battery
(*a*) general view; (*b*) plates

active material of the electrodes is reduced. The concentration of the electrolyte remains the same during discharge.

On discharge, the voltage rapidly falls from 1.4 V to 1.3 V, then more slowly to 1.1 V which is the final voltage. On charging, the voltage rapidly rises from 1.15 V to 1.75 V, and then, after some insignificant fall, increases slowly to 1.85 V. The internal resistance of alkaline storage cells is higher than that of acid storage cells, so they have a smaller overall efficiency, $\eta = 0.5\text{-}0.6$, and a higher resistance to short circuits. Alkaline storage cells offer a number of advantages over the lead-acid type, namely: greater mechanical strength, longer service life, and need less skilled attendance.

A *silver-zinc storage cell* consists of two plate groups located in a plastic container holding an electrolyte. The cell electrodes are porous plates; the positive plate is made of silver oxide (Ag_2O) and the negative, of zinc (Zn). The electrolyte is an aqueous solution of potassium hydrate (KOH) with a specific gravity of 1.4.

During discharge, the silver oxide changes to metallic silver, and the metallic zinc changes to zinc oxide. On charging, the reactions are reversed.

During charge, the voltage at first remains almost the same (1.65 V), then it rapidly rises to 1.9 V and more slowly to 2.1 V, at which the charging should be stopped. On discharge, the voltage drops off slowly from 1.75 V to 1.5 V and then to 1.25 V-1 V at the end of discharge, which is the final voltage.

The advantages of the silver-zinc cell are as follows:

(1) considerably greater capacity and power per unit mass;

(2) a stable discharge voltage of 1.5 V and ability to furnish very high currents for short discharge periods;

(3) a high overall efficiency of about 0.85.

2-19. Combinations of Chemical Current Sources

If the voltage and current necessary for operation exceed the respective quantities that can be supplied by one cell, several cells are combined to make up a battery.

The cells combined into a battery should have equal emfs, E_0, and equal internal resistances, r_0.

Cells are connected *in series* (Fig. 2-16) when the load current does not exceed the nominal current of a single cell, while the load voltage V is higher than the emf of the cell (E_0). In this case, the number of cells connected in series n is determined by the ratio $n \geqslant V/E_0$. The emfs of the component cells all point in the same direction, because the negative pole of a preceding cell is connected to the positive pole of the next, so the total emf of a battery is n times the emf of a single cell:

$$E = nE_0 \tag{2-56}$$

as is the internal resistance of the battery

$$r = nr_0 \tag{2-57}$$

The discharge current of a battery is equal to the discharge current of an individual cell.

Cells are connected *in parallel* (Fig. 2-17) where the load voltage V is equal to the emf of a single cell V_0 while the load current I exceeds greatly the discharge current of a cell I_d. In this case, the number of cells, m, connected in parallel is determined by the ratio $m > I/I_d$. Cells are said to be connected in parallel when their positive and negative poles are connected to the respective common terminals;

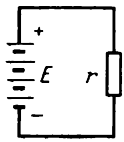


Fig. 2-16. Series connection of cells

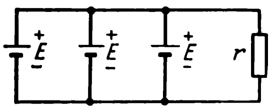


Fig. 2-17. Parallel connection of cells

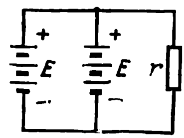


Fig. 2-18. Series-parallel connection of cells

thus, the emf of the battery E is equal to the emf E_0 of an individual cell:

$$E = E_0 \tag{2-58}$$

the internal resistance of the battery is

$$r = r_0/m \tag{2-59}$$

and the discharge current of the battery is m times the emf of a single cell:

$$I = I_d m \tag{2-60}$$

Cells are connected in *series-parallel* (Fig. 2-18) where the load voltage and current exceed the nominal voltage and current of a cell. The number of series-connected cells n and the number of parallel branches m are determined by the formulas given above.

Example 2-13. Find the characteristics of the storage battery required to supply 2.9 kW at a voltage of $V = 120$ V for emergency lighting; the emf of each cell is $E_0 = 2$ V and its discharge current is $I_d = 6$ A.

Solution.

The current for emergency lighting is

$$I = P/V = 2900/120 = \text{(approx.) } 24 \text{ A}$$

The number of series-connected cells is

$$n = V/E_0 = 120/2 = 60$$

The number of parallel branches is

$$m = I/I_d = 24/6 = 4$$

The number of cells in the battery is

$$nm = 60 \times 4 = 240$$

2-20. Nonlinear Circuits

An electric circuit whose resistance depends neither on current nor on voltage is called *linear*. If the resistance of any section in an electric circuit depends on current and voltage, the section and the circuit itself are termed *nonlinear*. Examples are incandescent lamps, vacuum-tube and semiconductor devices. The resistance of a nonlinear circuit varies, so its current is not proportional to its voltage, that is, Ohm's law cannot be used to solve the circuit. Nonlinear circuits are usually solved graphically.

The relationship between the current in and the voltage across a circuit, $I = f(V)$, is called its *volt-ampere characteristic*. For a linear element, the volt-ampere characteristic is a straight line, for example, *Oa* (Fig. 2-19), which passes through the origin. For a nonlinear element, it is no longer a straight line, but a curve. For example, curves *Ob* and *Oc* represent the volt-ampere characteristics of carbon and metal filament lamps, respectively.

To solve a simple (single-mesh) circuit having two nonlinear components (Fig. 2-20), we choose scales for its current and voltage and construct the following volt-ampere characteristics: $I_1 = f_1(V_1)$ for the element NE_1, and $I_2 = f_2(V_2)$ for the element NE_2, on the same coordinates. Combining the voltages V_1 and V_2, both corresponding to a particular current, we obtain a voltage V across the circuit, that is, a certain point on the $I = f(V)$ volt-ampere characteristic (Fig. 2-21). For example, the point A' on the volt-ampere characteristic for an arbitrary current I' is obtained by adding together abscissa $A'_0A'_1$ and abscissa $A'_0A'_2$.

When finding the circuit current at a specified voltage V, the voltage is plotted as abscissa (intercept OO'). The per-

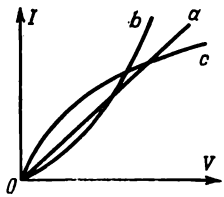


Fig. 2-19. Volt-ampere characteristics

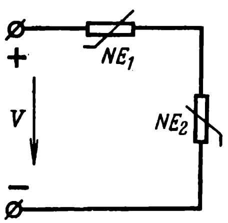


Fig. 2-20. Simple (single-loop) circuit with two nonlinear elements

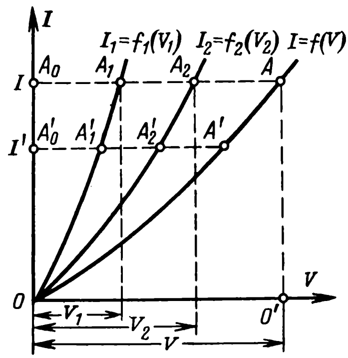


Fig. 2-21. Volt-ampere characteristics of a simple (single-loop) circuit

pendicular $O'A$ raised from the point O' until it intersects the $I = f(V)$ volt-ampere characteristic of the circuit locates the current I. The lines A_0A_1 and A_0A_2 drawn through the point A parallel to the x-axis make intercepts which represent voltages V_1 and V_2 across the nonlinear elements.

A circuit having parallel-connected nonlinear elements (Fig. 2-22) for a specified voltage across the circuit V_s is solved by finding the branch currents from the respective volt-ampere characteristics (Fig. 2-23). The branch voltages are the same, so, plotting the voltage as abscissa (intercept OO'), we find the currents I_1 and I_2 (ordinates $O'A_1$ and $O'A_2$). The total circuit current is the sum of the branch currents: $I = I_1 + I_2$.

If it is necessary to find the branch currents from a specified total current I, we construct the overall volt-ampere characteristic $I = f(V)$ by combining the ordinates of the

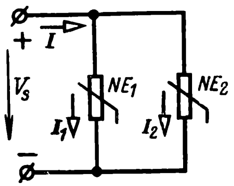


Fig. 2-22. Parallel connection of two nonlinear elements

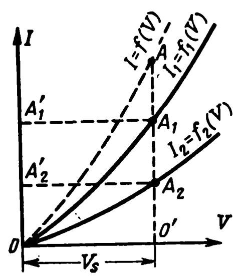


Fig. 2-23. Volt-ampere characteristics (for a parallel circuit)

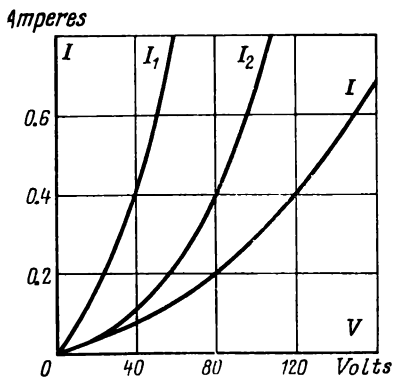


Fig. 2-24. Illustrating Example 2-14

branch volt-ampere characteristics for the same voltages (Fig. 2-23).

Assuming a specified current I (the point A on the overall volt-ampere characteristic), we determine the voltage $V_s = V_1 = V_2$ (the point O'), and the currents I_1 and I_2 (the points A_1 and A_2).

Example 2-14. Find the current in and the voltage across two series-connected nonlinear elements, if the supply voltage is $V_s = 120$ V.

Solution.

The volt-ampere characteristics are constructed by the data given in Table 2-4.

Then we find the voltages for the same currents 0.1, 0.2, 0.3, 0.4, 0.5 A for both elements, plot them as abscissas and construct the volt-ampere characteristics for the two

Table 2-4

V	V	0	20	40	60	80	100
I_1	A	0	0.16	0.4	0.8	—	—
I_2	A	0	0.04	0.12	0.22	0.40	0.65

elements. After that, we combine the individual characteristics to obtain the overall volt-ampere characteristic of the circuit (Fig. 2-24).

Using this characteristic, we find that for $V_s = 120$ V the current is $I = 0.4$ A. For this current, the voltages across the elements are $V_1 = 40$ V and $V_2 = 80$ V.

Chapter Three Electromagnetism

3-1. Magnetic Field of a Current

A conductor carrying an electric current is surrounded by a magnetic field which, as stated in Sec. 1-1, is a form of matter. It has experimentally been found that a magnetic field arises not only around or in conductors carrying a current, but also owing to the motion of any electrically charged particles or bodies, or variations in an electric field. For example, the magnetic field around permanent magnets owes its existence to molecular currents, resulting from the motion of electrons in their orbits and rotation about their axes. A magnetic field manifests itself through its action on moving charged particles, specifically, on an electric current in a wire (Sec. 3-5), and also on permanent or electric magnets.

A magnetic field can best be detected by a compass needle. The needle brought into the magnetic field of a current-

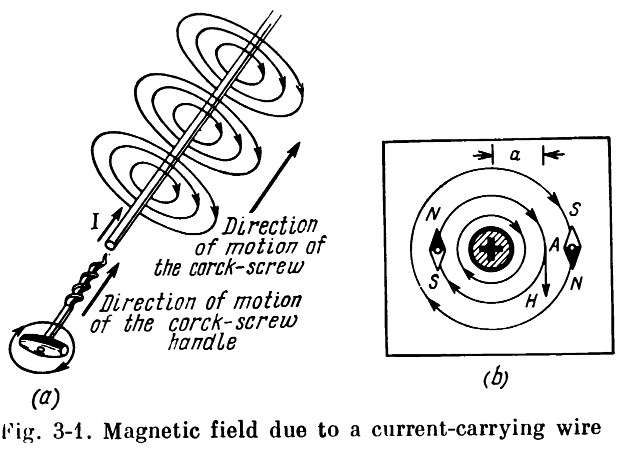


Fig. 3-1. Magnetic field due to a current-carrying wire

carrying wire tends, under the action of the field, to align itself at right angles to the axis of the wire (Fig. 3-1). The direction in which the "north" end of the compass needle points is taken as the direction of the magnetic field in the region where the needle is placed.

Graphically, a magnetic field is conventionally depicted by magnetic lines of force. Magnetic lines of force are drawn in such a way that the direction of a line tangent to a line of force at any point coincides with the direction of the field. If we draw magnetic lines of force through any unit area (1 m^2, 1 cm^2, at right angles to the direction of the lines), so that the number of lines is proportional to the field intensity at a given point, the magnetic field strength characterized by magnetic induction will be represented by the density of magnetic lines of force (see Sec. 3-4). The magnetic lines of force or lines of magnetic flux are always closed. For example, the magnetic lines of force of a straight current-carrying wire (Fig. 3-1*b*) have the form of concentric circles located in the planes passing at right angles to the wire axis.

The direction of the magnetic lines around a wire can be determined by what is known as the right-hand screw rule (Fig. 3-1*a*). *If a right-hand screw is moved progressively in the direction of the current, the rotation of its handle will give the direction of the magnetic lines of force around the wire.* Conversely, we can find the direction of a current, if we know the direction of the associated magnetic lines of force. Fig. 3-2*a* shows the magnetic lines of force around a circular current-carrying conductor, and Fig. 3-2*b* shows the lines around a current-carrying coil.

In the case of a ring or coil, it is more convenient to apply the right-hand screw rule as follows. *If the screw is rotated in the direction of the current in the ring or coil, the direction of its progressive motion will give the direction of the magnetic lines of force linked by the surface bounded by the current loop.*

Thus, the direction of a magnetic field depends on the direction of the associated current.

A magnetic field is called *uniform*, if it has the same direction and intensity at all points. Otherwise, it is termed *nonuniform*. Graphically, a uniform magnetic field is shown by parallel magnetic lines of force spaced an equal distance

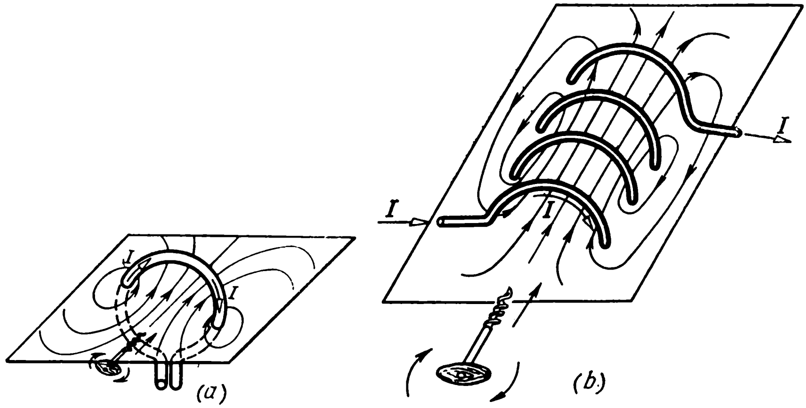


Fig. 3-2. Magnetic field (*a*) of a ring current and (*b*) due to a current-carrying coil

apart. For example, such a field can be obtained in the air gap between two flat parallel pole-pieces of a magnet or electromagnet.

The magnetic field of a straight current-conducting wire is symmetrical. This means that in a plane at right angles to the wire axis, all points equally distant from the wire are in identical physical conditions. Hence, the field strength is the same at all such points, or, to state this differently, at an arbitrary fixed point of the field at the distance r from the wire, the field strength remains the same, as the

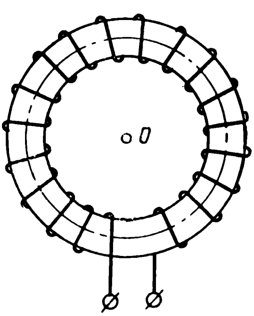


Fig. 3-3. Ring-shaped coil

wire is rotated about its axis. This is also true of any other plane parallel to the above one.

Another example of a symmetrical field is that around a ring-shaped coil whose turns are uniformly distributed along the core (Fig. 3-3). As directly follows from the condition of symmetry, the magnetic lines of force in the core are circles whose centres lie on a straight line perpendicular to the plane of the drawing and passing through the centre O of the coil. It is likewise obvious that the field strength is the same at all points on the same magnetic line of force.

3-2. Magnetomotive Force, Magnetic Field Strength

The electric current in and the magnetic field around a current-carrying conductor are two inseparable facets of a single entity known as an electromagnetic field. It is customary, however, to say that an electric current has the property of producing a *magnetic field.* This property or, rather, ability of an electric current is termed a *magnetizing force* (symbolized by the letter H) or a magnetomotive force, mmf, symbolized by the letter F.

Formally, an mmf produces a magnetic field (a magnetic flux, see Sec. 3-4) in much the same way as an emf gives rise to an electric current in an electric circuit.

In the International System of Units, an mmf is assumed to be numerically equal to the current producing the magnetic field, so the mmf of a current-carrying conductor is $F = I$. If a current has its path round a loop or in a coil with w turns, the mmf is equal to the product of the current and the number of turns, that is, $F = Iw$.

Naturally, an mmf, like a current, is measured in amperes*

$$[F] = [I] = \mathrm{A}$$

The direction of the mmf of a current-carrying coil or loop can conveniently be determined by the *right-hand coil rule*, which states: *If you grasp the coil with your right hand so that your fingers go around it in the direction of the current in the wires, then the extended thumb will point in the direction*

* In the US and UK literature of the subject, the measure of magnetomotive force is an ampere-turn.— *Translator's note.*

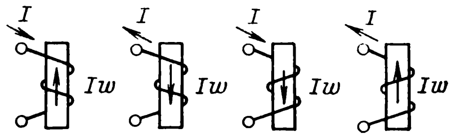


Fig. 3-4. MMFs of current-carrying coils

of the mmf. Figure 3-4 shows several current-carrying coils and the directions of their respective mmfs.

The mmf in a symmetrical field, say, one due to a ring-shaped coil, is uniformly distributed along a magnetic line of force. The fraction of the total mmf per unit length of a magnetic line of force is called the *magnetic field strength* or *magnetic intensity* H and is an important characteristic of magnetic fields.

The magnetic field strength at a given point is a function of the current in, and the geometry of, the conductor and is independent of the properties of the medium, if the latter is uniform. The magnetic field strength is a vector quantity. In isotropic media, that is, those having the same magnetic properties in any direction, the magnetic intensity vector has the same direction as the magnetic line of force at that point. In the International System of Units, the magnetic field strength or magnetic intensity is measured in amperes per metre*

$$[H] = [F/l] = \text{A/m}$$

Sometimes, use is made of the oersted, which is not an SI unit, defined as

$$1 \text{ oersted} = \text{(approx.) } 80 \text{ A/m} = 0.8 \text{ A/cm}$$

If a magnetic field is symmetrical, it is an easy matter to compute its strength. For example, the magnetic field strength at a point A which is a distant from the axis of a straight current-carrying conductor (Fig. 3-1b) will, in

* In the US and UK literature of the subject, the measure of magnetic field strength or magnetic intensity is an ampere-turn per metre.— *Translator's note.*

accordance with the above definition, be given by

$$H = I/l = I/2\pi a \tag{3-1}$$

where l is the length of a magnetic line of force of radius a.

For example, if a wire carries a current $I = 100$ A, the magnetic field strength within 10 cm (0.1 m) of the wire's axis will be

$$H = I/2\pi a = 100/(2 \times 3.14 \times 0.1)$$

$$= 100/0.628 = \text{(approx.) } 160 \text{ A/m}$$

By analogy with an electric potential difference, one often uses the concept of *magnetic potential difference* U_m.

The magnetic potential difference between two points in a uniform magnetic field, located on a common magnetic line of force, is given by the product of the magnetic field strength and the separation between the points:

$$V_m = Hl \tag{3-2a}$$

In a nonuniform magnetic field, the magnetic potential difference between two points in a field is the sum of the elementary magnetic potential differences $H\Delta l$ over elementary intervals Δl along the specified path between the points:

$$V_m = \Sigma H\Delta l \tag{3-2b}$$

In the International System of Units, the magnetic potential difference is measured in amperes*

$$[V_m] = [Hl] = (\text{A/m})\,\text{m} = \text{A}$$

The magnetic potential difference around an arbitrary closed contour gives the mmf. Thus, the mmf may be defined as the sum of the magnetic potential differences ($H\Delta l$) around a closed contour

$$F = \Sigma H\Delta l$$

3-3. Ampere's Circuital Law

The algebraic sum of the currents linked by a surface bounded by a closed contour is referred to as the total current.

---

* In the UK and US literature of the subject, the measure of magnetic potential difference is ampere-turn.— *Translator's note.*

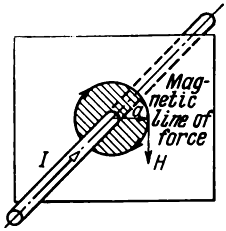


Fig. 3-5. Current-carrying wire passing through a surface perpendicular to the wire axis

Using Eq. (3-1), we may write for the magnetic field strength at a distance a (Fig. 3-1b) from the axis of a straight conductor carrying a current I

$$I = H \times 2\pi a = Hl \tag{3-3}$$

The product of the magnetic field strength H by the length of the magnetic line of force, $l = 2\pi a$, which bounds the closed contour, is the mmf, F_m. The surface bounded by the magnetic line of force (Fig. 3-5) in our case links the same current I, so the algebraic sum of currents is $\Sigma I = I$.

Equation (3-3) may be re-written as

$$\Sigma I = F_m \tag{3-4}$$

Thus, *the mmf around a closed contour is equal to the total current linked by the surface bounded by that contour.* This relation is known as Ampere's circuital law. We have derived it from a simple example.

If the magnetic field strength is not the same along a magnetic line of force, the mmf is found as the sum of the products Hl for the various sections, that is,

$$F_m = H_1 l_1 + H_2 l_2 + H_3 l_3 + \dots = \Sigma Hl$$

3-4. Magnetic Induction, Permeability, Magnetic Flux

When an unvarying current I traverses a coil of w turns, the magnetic field strength H remains unchanged. This observation also holds when an iron core is placed inside the coil. In the latter case, however, the field density inside the coil increases appreciably owing to the molecular currents of the core, that is, owing to the change in the state of the medium (the core) in which the magnetic field exists. The

density of a field at each point, adjusted for the effect of the medium, is characterized in terms of the *magnetic induction*, B.

Magnetic induction is a vector associated with the force that the magnetic field exerts on a current element (see Sec. 3-5). Thus, *magnetic induction is a force characteristic of a magnetic field.*

The magnetic induction vector points in the same direction as a line tangent to the magnetic lines of force. In isotropic media, this is the direction of the magnetic intensity vector. Also, because the density of magnetic lines of force is proportional to the field intensity characterized by the magnetic induction, it may be said that the magnetic lines of force are also lines of magnetic induction.

Magnetic induction and magnetic field strength are connected by a simple relation of the form

$$B = \mu_a H \tag{3-5}$$

where μ_a is the *absolute permeability* of the medium.

From a comparison of the magnetic field due to the current in a wire placed in a given medium and in a vacuum, it has been found that the properties of the medium affect the magnetic field strength. In a paramagnetic medium, the field is augmented; in a diamagnetic medium, it is weakened. Thus, the magnetic induction B is a function of the properties of the medium in which the field exists.

The absolute permeability of vacuum is termed the *magnetic constant* (its symbol is μ_0 or, sometimes, μ_v). In the International System of Units, it is defined as

$$\mu_0 = 4\pi 10^{-7} \text{ ohms} \times \text{s/m}$$

One ohm·second (Ωs) is called the henry which is the unit of inductance (see Sec. 3-16). Thus,

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m} = (\text{approx.})\, 125 \times 10^{-8} \text{ H/m} \tag{3-6}$$

The absolute permeability of materials (media) is compared with the magnetic constant. The ratio of the absolute permeability of a material to the magnetic constant is termed the *relative permeability* of that material, symbolized by the letter μ:

$$\mu = \mu_a/\mu_0 \tag{3-7}$$

The relative permeability of a material is a dimensionless quantity. For diamagnetic materials and media, it is less than unity. For example, for copper $\mu = 0.999995$, for paramagnetic materials $\mu > 1$, for air $\mu = 1.0000031$. For engineering purposes, the relative permeability of diamagnetic and paramagnetic materials and media is assumed to be unity.

The relative permeability of ferromagnetic materials (see Sec. 3-9) which play an exceptionally important part in electrical engineering, runs into tens of thousands and depends on the material, temperature, induction, and magnetic field strength.

The unit of magnetic induction, the tesla, can be deduced from Eq. (3-5)

$$[B] = [\mu_a H] = [\mu_0 \mu H] = (\mathrm{H/m})\,(\mathrm{A/m})$$
$$= \frac{\Omega \times \mathrm{s} \times \mathrm{A}}{\mathrm{m}^2} = \frac{\mathrm{V\,s}}{\mathrm{m}^2} = \mathrm{T}$$

The volt second, V s, is called the weber, Wb, the unit of magnetic flux. The weber per square metre, $\mathrm{Wb/m^2}$, is called the tesla, T. Thus, in the International System of Units, magnetic induction can be measured in webers per square metre or in teslas, $[B] = \mathrm{Wb/m^2} = \mathrm{T}$.

In the calculation of magnetic fields, use is often made of the gauss, G, which is not an SI unit:

$$1\ \mathrm{G} = 10^{-4}\ \mathrm{T} = 10^{-4}\ \mathrm{Wb/m^2}$$

The product of the magnetic induction B of a uniform field by the surface area S perpendicular to the magnetic induction vector is called the *magnetic flux*

$$\Phi = BS \tag{3-8}$$

As already noted, the SI unit of magnetic flux is the weber or volt second:

$$[\Phi] = [BS] = \frac{\mathrm{V\,s}}{\mathrm{m}^2}\,\mathrm{m}^2 = \mathrm{V\,s} = \mathrm{Wb}$$

There is a smaller unit of magnetic flux, not covered by the SI; it is called the maxwell, Mx

$$1\ \mathrm{Mx} = 10^{-8}\ \mathrm{Wb}$$

density of a field at each point, adjusted for the effect of the medium, is characterized in terms of the *magnetic induction*, B.

Magnetic induction is a vector associated with the force that the magnetic field exerts on a current element (see Sec. 3-5). Thus, *magnetic induction is a force characteristic of a magnetic field.*

The magnetic induction vector points in the same direction as a line tangent to the magnetic lines of force. In isotropic media, this is the direction of the magnetic intensity vector. Also, because the density of magnetic lines of force is proportional to the field intensity characterized by the magnetic induction, it may be said that the magnetic lines of force are also lines of magnetic induction.

Magnetic induction and magnetic field strength are connected by a simple relation of the form

$$B = \mu_a H \tag{3-5}$$

where μ_a is the *absolute permeability* of the medium.

From a comparison of the magnetic field due to the current in a wire placed in a given medium and in a vacuum, it has been found that the properties of the medium affect the magnetic field strength. In a paramagnetic medium, the field is augmented; in a diamagnetic medium, it is weakened. Thus, the magnetic induction B is a function of the properties of the medium in which the field exists.

The absolute permeability of vacuum is termed the *magnetic constant* (its symbol is μ_0 or, sometimes, μ_v). In the International System of Units, it is defined as

$$\mu_0 = 4\pi 10^{-7} \text{ ohms} \times \text{s/m}$$

One ohm·second (Ωs) is called the henry which is the unit of inductance (see Sec. 3-16). Thus,

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m} = (\text{approx.})\, 125 \times 10^{-8} \text{ H/m} \tag{3-6}$$

The absolute permeability of materials (media) is compared with the magnetic constant. The ratio of the absolute permeability of a material to the magnetic constant is termed the *relative permeability* of that material, symbolized by the letter μ:

$$\mu = \mu_a / \mu_0 \tag{3-7}$$

The relative permeability of a material is a dimensionless quantity. For diamagnetic materials and media, it is less than unity. For example, for copper $\mu = 0.999995$, for paramagnetic materials $\mu > 1$, for air $\mu = 1.0000031$. For engineering purposes, the relative permeability of diamagnetic and paramagnetic materials and media is assumed to be unity.

The relative permeability of ferromagnetic materials (see Sec. 3-9) which play an exceptionally important part in electrical engineering, runs into tens of thousands and depends on the material, temperature, induction, and magnetic field strength.

The unit of magnetic induction, the tesla, can be deduced from Eq. (3-5)

$$[B] = [\mu_a H] = [\mu_0 \mu H] = (\mathrm{H/m})\,(\mathrm{A/m})$$
$$= \frac{\Omega \times \mathrm{s} \times \mathrm{A}}{\mathrm{m}^2} = \frac{\mathrm{V\,s}}{\mathrm{m}^2} = \mathrm{T}$$

The volt second, V s, is called the weber, Wb, the unit of magnetic flux. The weber per square metre, $\mathrm{Wb/m^2}$, is called the tesla, T. Thus, in the International System of Units, magnetic induction can be measured in webers per square metre or in teslas, $[B] = \mathrm{Wb/m^2} = \mathrm{T}$.

In the calculation of magnetic fields, use is often made of the gauss, G, which is not an SI unit:

$$1\ \mathrm{G} = 10^{-4}\ \mathrm{T} = 10^{-4}\ \mathrm{Wb/m^2}$$

The product of the magnetic induction B of a uniform field by the surface area S perpendicular to the magnetic induction vector is called the *magnetic flux*

$$\Phi = BS \tag{3-8}$$

As already noted, the SI unit of magnetic flux is the weber or volt second:

$$[\Phi] = [BS] = \frac{\mathrm{V\,s}}{\mathrm{m}^2}\,\mathrm{m}^2 = \mathrm{V\,s} = \mathrm{Wb}$$

There is a smaller unit of magnetic flux, not covered by the SI; it is called the maxwell, Mx

$$1\ \mathrm{Mx} = 10^{-8}\ \mathrm{Wb}$$

The magnetic flux is an important physical quantity. As will be shown shortly, the work done by a conductor carrying a current I in cutting a magnetic flux Φ is equal to the product of the magnetic flux by the current, that is, $W = \Phi I$. As is shown in Sec. 3-12, the rate of change of the magnetic flux linking a current-carrying loop is equal to the emf induced in that loop.

The magnetic field strength in a uniform medium surrounding a straight current-carrying conductor (Fig. 3-1b) is given by Eq. (3-1).

The product of the magnetic field strength by the absolute permeability of the medium gives the magnetic induction

$$B = \mu_a H = \mu\mu_0 (I/2\pi a) = 4\pi\mu (I/2\pi a) \times 10^{-7} \qquad (3\text{-}9)$$

where the current I is in amperes, the distance a is in metres, and the magnetic induction B is in teslas.

If a current-carrying conductor is placed in a non-ferromagnetic medium, then, on setting $\mu = 1$, we get

$$B = 4\pi (I/2\pi a) \times 10^{-7} = (2I/a) \times 10^{-7} \qquad (3\text{-}9a)$$

This equation holds for any value of a greater than the wire radius and for a wire of infinite length. However, it is also applicable to a finite conductor, if the distance a is a fraction of the wire length.

3-5. Electromagnetic Force

(a) A Straight Current-Carrying Wire in a Magnetic Field

If we place a straight conductor carrying a current I in a magnetic field at right angles to its lines of force (Fig. 3-6a), the field will exert on it an electromagnetic force F. This force is proportional to the current I, the active length l of the conductor (the part of the conductor which lies inside the field), and the magnetic induction B

$$F = IBl \qquad (3\text{-}10)$$

If the current is in amperes, the magnetic induction in teslas, and the length in metres, then the electromagnetic force will be in newtons

$$[F] = [IBl] = \mathrm{A} \times \mathrm{T} \times \mathrm{m} = \mathrm{A}\,\frac{\mathrm{V\,s}}{\mathrm{m}^2}\,\mathrm{m} = \frac{\mathrm{A} \times \mathrm{V\,s}}{\mathrm{m}} = \mathrm{J/m} = \mathrm{N}$$

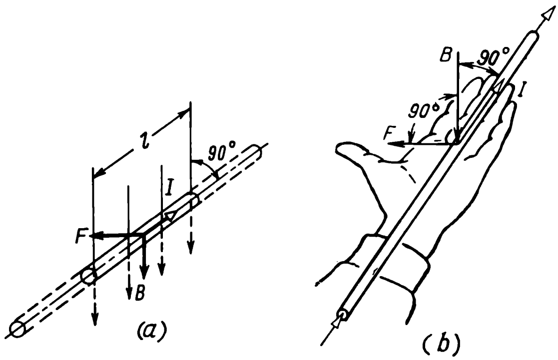


Fig. 3-6. (*a*) Current-carrying wire in a magnetic field; (*b*) the left-hand (wire) rule

The direction of the electromagnetic force F (Fig. 3-6*b*) can conveniently be determined by the *left-hand rule*: *If you place your left palm so that the magnetic induction vector enters it, and extend the four fingers in the direction of the current flow, then the thumb extended at right angles will give the direction of the electromagnetic force.*

If we move a current-carrying conductor through a uniform magnetic field so as to vary the angle α between the direction of the conductor and that of the field, while holding I, B and l constant, the force exerted on the conductor will vary in proportion to $\sin\alpha$. When the conductor is parallel to the magnetic lines of force, the electromagnetic force is zero.

Thus, in the general case

$$F = IBl \sin\alpha \qquad (3\text{-}10a)$$

The force that a magnetic field exerts on current-carrying conductors is widely utilized in a variety of electromagnetic mechanisms, notably electric motors.

Example 3-1. Given: A wire with an active length of 20 cm (0.2 m), carrying a current of 300 A, placed in a uniform magnetic field of 1.2 T induction.

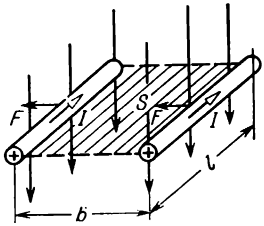


Fig. 3-7. Wire moved through a distance b in a magnetic field

To find: The electromagnetic force acting on the conductor when it lies in a plane perpendicular to the field.

Solution.

$$F = BIl = 1.2 \times 300 \times 0.2 = 72 \text{ N}$$

Suppose that a straight conductor carrying a current I is placed in a uniform magnetic field in a plane perpendicular to the field (Fig. 3-7). If the electromagnetic force causes this conductor to move in this plane in a direction normal to its axis for a distance b, the conductor will cut a magnetic flux given by $\Phi = BS = Blb$. As a result, the electromagnetic force will have done mechanical work given by

$$W = Fb = IBlb = I\Phi \tag{3-11}$$

Or, to state this in words, the mechanical work done in moving a current-carrying conductor in a magnetic field is equal to the product of the current in the wire by the magnetic flux it cuts. The energy expended in doing this work is supplied by an external power source.

Example 3-2. Given: A uniform magnetic field of 1.5 T induction and a wire 30 cm (0.3 m) long, carrying a current of 200 A.

To find: The work done in moving the wire for a distance of 20 cm (0.2 m) in a plane perpendicular to the field.

Solution.

The magnetic flux cut by the wire is

$$\Phi = BS = 1.5 \times 0.3 \times 0.2 = 0.09 \text{ Wb}$$

The work done in moving the wire is

$$W = \Phi I = 0.09 \times 200 = 18 \text{ J}$$

(b) A Current-Carrying Loop in a Magnetic Field

Let us place the sides of a current-carrying rectangular coil or loop (Fig. 3-8*a*), which are at right angles to the plane of the drawing, in a uniform field. We shall note that

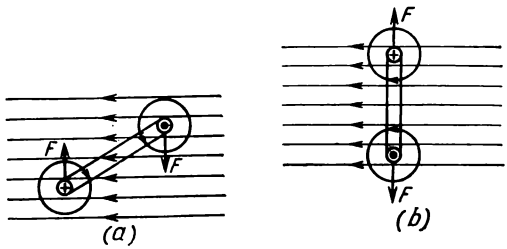


Fig. 3-8. Electromagnetic forces acting on current-carrying loop

the electromagnetic force F gives rise to a turning moment, or torque. This torque will cause the loop to take up a position where the applied forces will balance out (Fig. 3-8*b*), but the loop will link a maximum magnetic flux. Therefore, we can conclude that *when a current-carrying loop is placed in a magnetic field, it tends, due to the action of electromagnetic forces, to take up a position where it links a maximum magnetic flux.*

(c) An Electron Moving in a Magnetic Field

The electromagnetic force that a magnetic field exerts on a current-carrying wire l long (see Fig. 3-6) is

$$F = IBl$$

This force may be regarded as the sum of the forces exerted on the free electrons of the wire that produce the electric current.

Let the number of free electrons in a wire l long be N. Then the force acting on an electron is

$$F_0 = F/N \tag{3-12}$$

If we designate the total charge carried by free electrons as $Q = Ne$ and their average velocity as $v = l/t$, then the

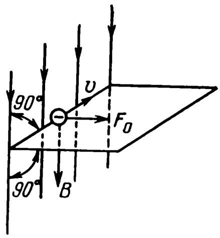


Fig. 3-9. Direction of the electromagnetic force acting on a moving electron

electromagnetic force exerted on an electron moving at right angles to the field will be

$$F_0 = F/N = (Q/t)\,Bl\,(1/N) = (Ne/t)\,Bl\,(1/N) = Bev \tag{3-13}$$

The direction of this force (Fig. 3-9) can be determined by the left-hand rule, with the four fingers extended against the motion of electron.

3-6. Interaction of Parallel Current-Carrying Conductors

Each of two parallel conductors carrying currents I_1 and I_2 (Fig. 3-10) gives rise to a magnetic field of its own. As a result, the first wire found in the magnetic field due to the current I_2 is acted upon by a force F_1, and the second in the magnetic field due to the current I_1, by a force F_2.

If the length of the parallel wires is considerably greater than the spacing a between them, then, by Eq. (3-9), the magnetic induction of each field within a distance a of the wires is

$$B_1 = \mu_a\,(I_1/2\pi a)$$

$$B_2 = \mu_a\,(I_2/2\pi a) \tag{3-14}$$

The vectors B_1 and B_2 are perpendicular to the plane passing through the wire axes, and their direction is given by the corkscrew rule.

By Eq. (3-10), the force exerted on the first wire is

$$F_1 = I_1B_2l = \mu_a\,(I_1I_2/2\pi a)\,l \tag{3-15}$$

Similarly, the force exerted on the second wire is

$$F_2 = I_2B_1l = \mu_a\,(I_1I_2/2\pi a)\,l \tag{3-16}$$

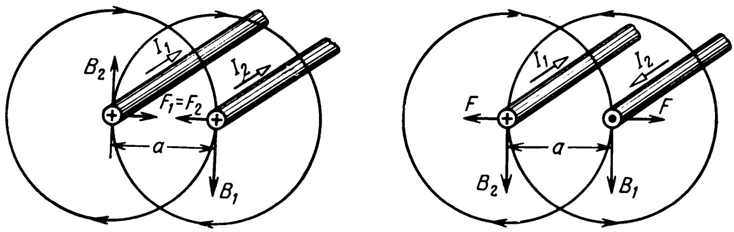


Fig. 3-10. Electromagnetic (electrodynamic) forces acting between current-carrying conductors

It follows from the last two equations that the forces exerted on the wires are equal, that is, $F_1 = F_2$. They are called *electrodynamic* forces.

When wires are placed in a vacuum or air, μ_a in Eqs. (3-15) and (3-16) is replaced by μ_0, because $\mu = 1$ and $\mu_a = \mu\mu_0 = \mu_0$.

3-7. The Magnetic Field due to a Current-Carrying Coil

The field can be concentrated in a particular region of space by passing a current through a coil. A further increase in the magnetic induction of the field can be achieved by increasing the number of turns in the coil and by fitting inside it an iron core whose molecular currents will produce a field of their own, thereby enhancing the main field set up by the coil turns.

Suppose we have a ring-shaped coil (Fig. 3-11) having w turns uniformly distributed along a non-magnetic core. The mean magnetic line of force coincides with a circle of radius R. This circle bounds a surface (or contour) which links a total current, $\Sigma I = Iw$.

Owing to its symmetry, the magnetic field strength H is the same at any point on the mean magnetic line of force, so the mmf is

$$F_m = Hl = H \times 2\pi R$$

By the Ampere circuital law,

$$Iw = Hl \qquad (3\text{-}17)$$

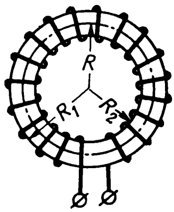


Fig. 3-11. Ring-shaped coil

whence the magnetic field strength along the mean magnetic line of force coincident with the axial line of the ring-shaped coil is

$$H = Iw/l \tag{3-18}$$

and the magnetic induction in teslas is

$$B = \mu_a H = \mu_a (Iw/l) = 125\mu(Iw/l) \times 10^{-8} \tag{3-19}$$

If $R_1 - R_2 \ll R_1$, the magnetic induction of the axial line of the coil may, with sufficient accuracy, be deemed equal to its average value, so the magnetic flux through the cross-sectional area of the coil is

$$\Phi = BS = \mu_a IwS/l \tag{3-20}$$

Equation (3-20) may be re-cast as follows

$$\Phi = Iw/[l/(\mu_a S)] = F_m/R_m \tag{3-21}$$

where Φ is the magnetic flux, F_m is the mmf, and $R_m = l/(\mu_a S)$ is the reluctance of the magnetic circuit (core) of the coil. Here, the reluctance is an attribute of a magnetic circuit corresponding to the resistance of an electric circuit. It is easy to recognize that Eq. (3-21) looks like Ohm's law for an electric circuit; it is known as *Bosanquet's law*.

A cylindrical coil (Fig. 3-12) may be regarded as part of a ring-shaped coil of a large radius, whose winding is positioned only on some part of the core with a length equal to the coil length. The magnetic field strength and magnetic induction on the axial line at the centre of a cylindrical coil can be found by Eqs. (3-18) and (3-19). However, these equations are approximate and hold only for coils for which $l \gg d$ (see Fig. 3-12).

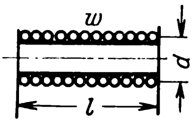


Fig. 3-12. Cylindrical coil

Example 3-5: Given: A cylindrical coil with a nonferromagnetic core ($\mu = 1$), 30 cm (0.3 m) long and 5 cm (0.05 m) in diameter, carrying 2000 turns.

To find: The magnetic flux of the coil for a current of 5 A.

Solution.

The magnetic flux of the coil is

$$\Phi = 4\pi\mu\,(IwS/l) \times 10^{-7} = 4\pi \times 1 \frac{5 \times 2000 \times 3.14 \times 0.05^2}{0.3 \times 4}$$

$$\times 10^{-7} = 8.22 \times 10^{-5}\ \text{Wb}$$

3-8. Ferromagnetics. Magnetization and Reversal of Magnetization

Ferromagnetics is a term applied to materials having high permeability. They include steel, iron, nickel, cobalt, their alloys, and some other metals.

The magnetic properties of a material depend on the magnetic properties of the elementary carriers of magnetism, that is, the electrons moving inside the atoms, and also on the interaction between their groups.

Moving in orbits around the atomic nucleus, the electrons give rise to *elementary currents* or *magnetic dipoles* which are characterized by the *magnetic dipole moment*, m. The magnetic dipole moment is given by the product of the elementary current i and the elementary surface S (Fig. 3-13) bounded by an elementary contour

$$m = iS$$

The vector m is perpendicular to the surface S, and its direction is determined by the corkscrew rule. The magnetic moment of a body is the vectorial sum of all of its individual magnetic dipole moments.

In addition to the magnetic dipole moment, there is also the magnetic moment arising from the spin of the electron

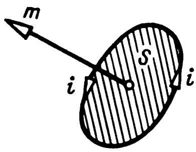


Fig. 3-13. Magnetic dipole moment due to an elementary current

about its axis. It is primarily the *electron-spin magnetic moment* that is responsible for ferromagnetism.

Present-day theory holds that ferromagnetic materials are composed of many small regions, or domains (10^{-2} to 10^{-6} cm^3 in size), each of which shows a spontaneous magnetization. When no external field is applied to a ferromagnetic material, the domains orient themselves in all likely directions, leaving the sample with zero resultant magnetization. When an external field is applied, such as one due to a current-carrying coil, the domains are rearranged so as to produce a net magnetization along the applied field, while the domains whose magnetic moments are directed approximately with the field grow in size. The result is the magnetization of the sample.

As the applied field is raised, an instant is finally achieved when all domains are arranged in the direction of the field and cease growing in size. The sample is said to have achieved a maximum state of magnetization; this condition is called *magnetic saturation.*

A magnetic circuit made up predominantly of ferromagnetic materials is capable of producing a large magnetic induction at a relatively small mmf.

In his experiments in 1872, Professor Stoletov of Moscow University placed an iron core in a current-carrying coil and measured the resultant magnetic induction B for various values of the magnetic field strength H; in his experiment he derived what is today known as the initial magnetization curve, $B = f(H)$, (Fig. 3-14).

The magnetization curve has three distinct regions, namely (1) a linear portion Oa which indicates that at first the magnetization builds up at a high rate and almost in proportion to the magnetic field strength; (2) a kink ab, within which the build-up of magnetic induction slows down, and (3) a linear portion past the kink ab, where B is a linear

function of H, although its rise is slow because of magnetic saturation.

The $B = f(H)$ curve shows that the permeability $\mu_a = B/H$ of a ferromagnetic material is a varying quantity and depends on the magnetic field strength.

In a.c. circuits, ferromagnetic materials undergo reversals of (or cyclic) magnetization.

A. G. Stoletov (1839-1896)

As the magnetizing current and, as a consequence, the magnetic field strength H is increased, the magnetic induction also increases until it attains a maximum value $+B_m$ (Fig. 3-15). As H is brought down, B also decreases, but more slowly and it has greater values for the same values of H than on the rising portion of the curve (region AC). When $H = 0$, the magnetic induction is still nonzero, and the remainder is called the *residual* induction, or remanence, B_r (the region OC in the plot of Fig. 3-15).

It follows from the foregoing that the magnetic induction of a sample depends not only on the magnetic field strength, but also on the previous history of the ferromagnetic material. The lag with which the magnetic induction follows changes in the magnetic field strength is called *magnetic hysteresis*. It may be visualized to arise from something like viscous friction as the magnetic moments of the domains change their orientation.

Reversals of the magnetizing current are accompanied by those of the magnetic field strength. The reverse H needed to reduce B_r to zero is called the *coercive force*, H_c (region OD).

As we keep increasing H in the reverse direction, the magnetic induction attains a negative maximum value, $-B_m$. If, now, we reduce H to zero, B will drop off to its residual value (region OF). If we reverse H again and bring it up to a maximum, the magnetic induction will reach the value designated $+B_m$ in the graph.

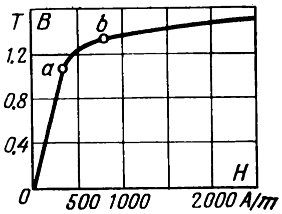


Fig. 3-14. Initial magnetization curve of steel

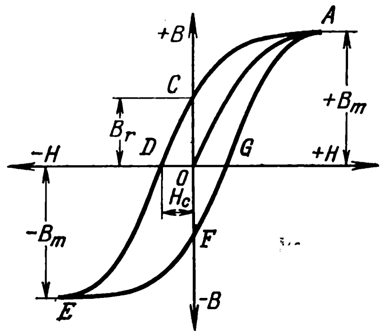


Fig. 3-15. Hysteresis loop

Thus, reversals of magnetization (or cyclic magnetization) of a ferromagnetic sample can graphically be depicted by a closed curve $ACDEFGA$, known as the *symmetric hysteresis loop*. The maximum hysteresis loop attainable by a given material is called its *major cyclic hysteresis loop*.

If we plot several symmetric hysteresis loops for a given ferromagnetic material at different values of B_m (Fig. 3-16) and join the tips of the loops, we shall obtain what is known as the *principal* or *major magnetization curve* which closely follows the initial magnetization curve.

The cyclic magnetization of steel raises its temperature, because hysteresis entails a loss of energy. The area of the

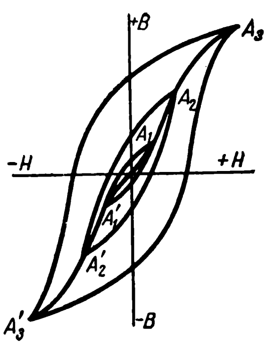


Fig. 3-16. Three hysteresis loops and the principal magnetization curves of steel

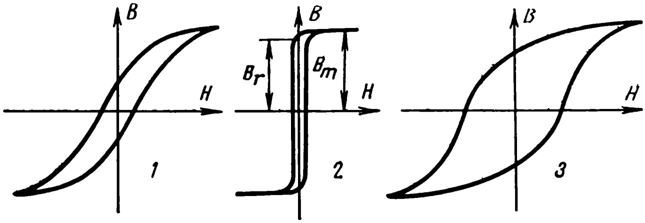


Fig. 3-17. Hysteresis loops for various materials
1—magnetically soft material (electrical-sheet steel); 2—magnetically soft material (Permalloy, square B-H loop); 3—magnetically hard material

hysteresis loop is proportional to the energy expended to carry out one cycle of magnetization.

The power lost due to hysteresis per unit mass of the material (specific hysteresis loss) depends on the quality of the material, magnetic induction, and the cycles of magnetization per second or, which is the same, the frequency f of the alternating current traversing the coil of the electromagnet.

The principal magnetization curve and the hysteresis loop of a sample characterize its properties. Figure 3-17 shows hysteresis loops for magnetically soft steel, Permalloy, and magnetically hard steel.

3-9. Ferromagnetic Materials

(a) Soft Magnetic Materials

Magnetically soft materials have high permeability, low coercive force (less than 400 A/m) and low specific hysteresis loss. Among them are ingot iron, Permalloys, electrical-sheet steels, and ferrites. They go to make magnetic circuits (cores) for d c. and a.c. applications.

Ingot iron (max 0.04% carbon) has high saturation induction (up to 2.2 T), high permeability (μ = 3500 to 7000), and low coercive force (H_c = 50 to 100 A/m). It is used in the manufacture of cores for d.c. magnetic fields.

Electrical-sheet steels are alloys of iron and silicon (1 to 4% Si). The silicon improves the properties of the iron. Among other things, it increases permeability, reduces coercive force and hysteresis loss and, what is most impor-

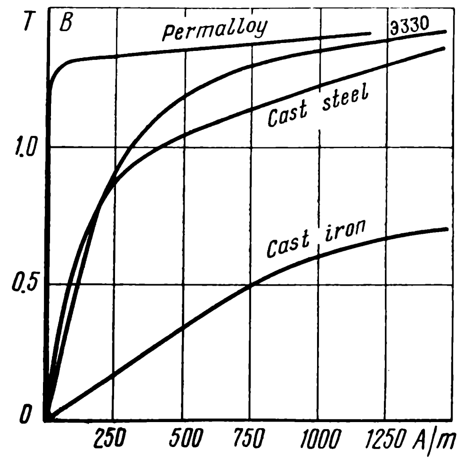


Fig. 3-18. Magnetization curve of several ferromagnetic materials

tant, raises the resistivity and, as a consequence, brings down eddy currents (Sec. 3-15) together with the associated losses.

Low-silicon steels have a lower permeability, a higher saturation induction, and a higher specific hysteresis loss. They are used in d.c. and low-frequency a.c. circuits operating at high magnetic induction.

High-silicon steels (2.8-4.8% Si) are used in circuits operating at power and high frequencies, where it is important to ensure low hysteresis and eddy-current losses or high permeability at low and medium field strengths.

In the Soviet Union, electrical-sheet steels are marked according to USSR State Standard GOST 802-58. The designation consists of the Russian letter "Э" (for "electric") and appropriate numerals. The first digit gives silicon content in per cent, the second defines the electromagnetic properties, and the third is "0" which indicates that the steel is cold-rolled.

The principal magnetization curve for Э330 electrical-sheet steel is shown in Fig. 3-18.

Permalloys are alloys of iron, nickel and some other elements. They have high permeability in weak magnetic

fields. They are classed into high-nickel (70-80% Ni) and low-nickel (40-50% Ni) grades.

The magnetic properties of Permalloys strongly depend on their nickel content and the manufacturing process used.

According to Soviet convention, the Russian letter "П" in the designation of a Permalloy indicates that it has a rectangular or, as is often said, square hysteresis (or *B-H*) loop (Fig. 3-17,2). The ratio of the residual induction B_r to the maximum induction B_m (B_r/B_m) is called the *squareness ratio of the material.* It may be as high as 0.95 to 0.99.

Ferrites can be of two types, cermet ferrites and oxide ferrites (or oxifers). The former are produced from metal-ceramic mixtures in a finely divided (powdered) state by pressing to shape and sintering. The latter are manufactured by the thermal decomposition of nickel and zinc salts. The two types are very close in magnetic properties, although within each type it is possible to fabricate various grades (magnetically hard, magnetically soft, with a square *B-H* loop, and so on), to suit a wide variety of applications.

Micropowder magnets (or powdered-dust cores) are materials prepared from fine particles of a ferromagnetic material (metal) bonded together by a dielectric such as PVC or polyethylene. The mixture is moulded, subjected to pressure, and baked.

Ferrites and powdered-dust cores are widely used in transformers, wire and radio communication equipment, computers, automatic control systems, etc. Especially wide use is made of ring-shaped square *B-H* loop cores in the memory (storage) units of computers. Such cores have the property to be magnetized to saturation by a current pulse and to store the residual magnetization for a long time.

(b) Hard Magnetic Materials

Magnetically hard materials have high coercive force and high remanence, so they make good permanent magnets in a variety of designs and applications. These materials include carbon steels, tungsten steels, chromium steels and cobalt steels which have a coercive force of 5000 to 13,000 A/m and a remanence of 0.7 to 1 T. They can readily be forged, rolled, and machined.

Hard magnetic materials also include alloys containing various amounts of iron, aluminium, nickel, silicon and cobalt, and known under various names, such as *Alni*, *Alnisi*, *Alnico* and *Magnico*. They have better magnetic properties than the materials listed in the previous paragraph. Their coercive force is from 20,000 to 60,000 A/m and the remanence from 0.2 to 2.25 T. These alloys are fabricated into magnets by casting, and the castings thus produced need then only grinding.

Powdered-dust magnets are fabricated by sintering fine particles of Alni and Alnico.

3-10. The Magnetic Circuit and Its Designing

A magnetic circuit is an assemblage of ferromagnetic cores with or without an air gap, which provides a closed path for a magnetic flux. The use of ferromagnetic materials stems from the desire to minimize the reluctance of the circuit, Eq. (3-21), to a value such that the desired magnetic induction or magnetic flux can be produced with a minimum mmf.

An elementary magnetic circuit is the core of a ring-shaped coil (Fig. 3-11). Magnetic circuits may be single-path and multi-path. In the latter, the various sections may differ in materials.

The design (or solution) of a magnetic circuit reduces to finding the mmf from the specified magnetic flux, circuit dimensions and materials.

More specifically, one divides the magnetic circuit into sections, so that each has a constant cross-section and a uniform field throughout, assigns the sections suitable indexes, say, l_1, l_2 and so on, finds the magnetic induction $B = \Phi/S$ for each section and, using the magnetization curves of Fig. 3-18, determines the respective magnetic field strengths. The magnetic field strength in an air gap or a nonferromagnetic material is

$$H_0 = B_0/\mu_0 = (\text{approx.})\, 0.8 \times 10^6 B_0 \qquad (3\text{-}22)$$

where H_0 is in amperes per metre and B_0 in teslas, or

$$H_0 = 0.8 B_0$$

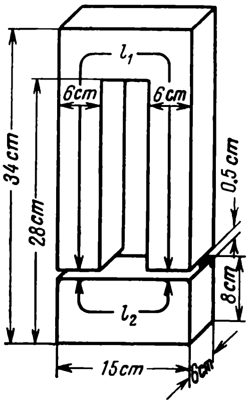


Fig. 3-19. Illustrating Example 3-6

if H_0 is in amperes per centimetre and B_0 in gauss.

By the Ampere circuital law, the sum of the magnetic field strengths in the individual sections gives the total current

$$H_1 l_1 + H_2 l_2 + H_0 l_0 + \ldots = Iw$$

Example 3-6. Determine the number of turns for the core of Fig. 3-19 required to obtain a magnetic flux of $47 \times \times 10^{-4}$ Wb if the coil is traversed by a current of 25 A. The top part of the core is Э330 electrical-sheet steel and the bottom part is a cast steel. The Э330 steel core is in three sections, namely $l_1 = 54$ cm (0.54 m) long and $S_1 = = 36$ cm^2 (0.0036 m^2) in cross-section, the second $l_2 = = 17$ cm (0.17 m) long and $S_2 = 48$ cm^2 (0.0048 m^2) in cross-section, and the third (the air gap) $l_0 = 0.5 \times 2 = = 1$ cm (0.01 m) long and $S_0 = 36$ cm^2 (0.0036 m^2) in cross-section.

Solution.

The magnetic induction in the first, second and third sections of the magnetic circuit are

$$B_1 = \Phi/S_1 = 47 \times 10^{-4} / 36 \times 10^{-4} = 1.3 \text{ T}$$

$$B_2 = \Phi/S_2 = 47 \times 10^{-4} / 48 \times 10^{-4} = 0.98 \text{ T}$$

$$B_0 = \Phi/S_0 = 47 \times 10^{-4} / 36 \times 10^{-4} = 1.3 \text{ T}$$

From the magnetization curve for Э330 electrical-sheet steel (see Fig. 3-18) we find that an induction of 1.3 T corresponds to a magnetic field strength of 750 A/m.

The magnetic potential difference across the first section is

$$V_{m1} = H_1 l_1 = 750 \times 0.54 = 405 \text{ A}$$

For the second section (see Fig. 3-18),

$$H_2 = 400 \text{ A/m}$$

and the magnetic potential difference is

$$V_{m2} = H_2 l_2 = 400 \times 0.17 = 68 \text{ A}$$

The magnetic field strength in the air gap is

$$H_0 = 0.8 \times 10^6 B_0 = 0.8 \times 10^6 \times 1.3$$
$$= 1.04 \times 10^6 \text{ A/m}$$

and the magnetic potential difference across the air gap is

$$V_{m0} = H_0 l_0 = 1.04 \times 10^6 \times 0.01 = 10{,}400 \text{ A}$$

Hence, the mmf is

$$F_m = V_{m1} + V_{m2} + V_{m0} = 405 + 68 + 10{,}400$$
$$= 10{,}873 \text{ A}$$

The number of turns in the coil must be

$$w = F_m/I = 10{,}873/25 = 435 \text{ turns}$$

3-11. Electromagnets

If we place an iron slug (an open magnetic circuit) near one end of a current-carrying coil as shown in Fig. 3-20, the electromagnetic forces will pull the slug inside the coil where it will take up a position for which the magnetic field is a maximum.

The above property is utilized in devices known as *electromagnets*.

An electromagnet consists of a current-carrying coil called a solenoid, and a magnetic circuit made up of a movable part, or armature (at *2* in Fig. 3-21), and a stationary part (at *1* in Fig. 3-21). When the solenoid is energized, the stationary part attracts the armature, the force

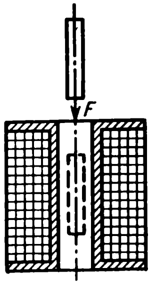


Fig. 3-20. Electromagnet with an open magnetic circuit

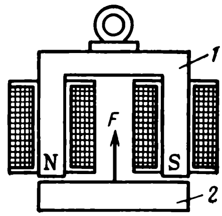


Fig. 3-21. Electromagnet with a closed magnetic circuit

of attraction being

$$F = \text{(approx.) } 4 \times 10^5 B^2 S \tag{3-23}$$

where F is in newtons, B is in teslas, and S (the cross-sectional area of the pole pieces) is in square metres.

If the magnetic circuit of a solenoid is unsaturated, it is possible to vary the magnetic induction and, as a consequence, the pull by varying the current traversing the solenoid.

Electromagnets have found many uses, for example to clamp steel workpieces in machine-tools, in automatic circuit breakers, relays, brakes, and so on.

Example 3-7. Given: The magnetic induction $B = 1.2$ T; the cross-sectional area of the pole-pieces $S = 200$ cm^2 (0.02 m^2).

To find: The pull of the electromagnet.

Solution. The pull of any electromagnet is given by

$$F = 4 \times 10^5 B^2 S$$

On substituting the specified numerical values in the above general equation, we obtain

$$F = \text{(approx.) } 4 \times 10^5 \times 1.2^2 \times 0.02 = 1.15 \times 10^4 \text{ N}$$

3-12. Electromagnetic Induction

(a) The EMF Induced in a Conductor

When a wire is moved with a velocity v, free electrons and positive ions of its material are moving with the same velocity. If the wire is moved in a uniform field at right angles to the magnetic lines of force (Fig. 3-22), every charged particle is acted upon by an electromagnetic force (see Sec. 3-5) whose direction is given by the left-hand rule. This force causes the electrons to move to one end of the wire, thereby building up a negative charge at that end and leaving a deficiency of electrons, or a positive charge, at the other. This separation of charges ceases as soon as the electromagnetic force is balanced by the electric force of attraction between the dissimilar charges. This chain of events gives rise to an emf in the wire, referred to as the *emf of electromagnetic induction*, and the process itself is called *electromagnetic induction*. It was discovered by Michael Faraday of England in 1831.

The voltage V between the ends of an open circuit is equal to the emf of electromagnetic induction E. Thus, according to Eq. (1-30), we have

$$E = \mathscr{E} l$$

Because, however, $\mathscr{E} = F_0/e$, and the force acting on an electron, according to Eq. (3-6), is $F_0 = Bve$, we have

$$E = Blv \tag{3-24}$$

Or, we may state, *the emf of electromagnetic induction induced in a wire is proportional to the magnetic induction of the field in which the wire is moving, the length and velocity of the wire in the direction normal to the magnetic lines of force*. This is one of the many statements of the law of electromagnetic induction.

The direction of the induced emf is given by the right-hand (wire) rule: *If you extend your right-hand palm so that the magnetic lines of force enter it and extend your thumb so that it points in the direction of the motion of the wire, then the four stretched fingers will indicate the direction of the induced emf* (Fig. 3-23).

If the wire is moved in a plane making an angle α with the magnetic induction vector, the emf will only be contri-

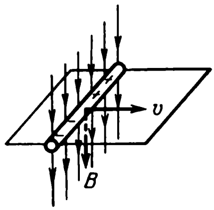


Fig. 3-22. Motion of a wire in a magnetic field

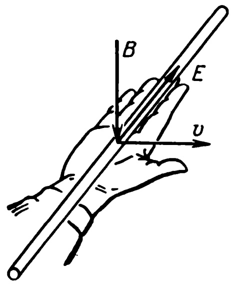


Fig. 3-23. Right-hand (wire) rule

buted by the velocity component perpendicular to the magnetic induction vector, or mathematically, $v_p = v \sin \alpha$, and the induced emf will be

$$E = Blv_p = Blv \sin \alpha \tag{3-25}$$

If the wire is moved in a plane normal to the magnetic lines of force with a velocity $v = \Delta b/\Delta t$, the induced emf will be

$$E = Blv = Bl\,(\Delta b/\Delta t)$$

Recalling that the product of magnetic induction B and an element of surface $\Delta S = l\Delta b$ gives the magnetic flux $\Delta\Phi = B\Delta S$ cut by the wire in time Δt, the induced emf will be

$$E = Bl\,(\Delta b/\Delta t) = \Delta\Phi/\Delta t \tag{3-26}$$

Or, to state this in words, *the induced emf is equal to the rate at which the wire cuts the magnetic flux.*

(b) The EMF Induced in a Loop

If a loop (Fig. 3-24) is moved in a nonuniform field in a plane normal to the magnetic lines of force (labelled by crosses in Fig. 3-24) in the direction marked by the arrow, this motion will induce emfs e_1 and e_2 in the loop sides *1* and *2*. The directions of these emfs, as found by the right-hand rule, are indicated by arrows. No emfs are induced in the loop sides *3* and *4*, because they do not cut the magnetic flux.

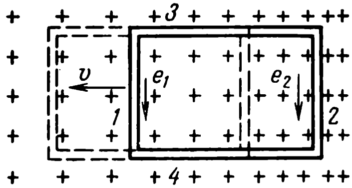


Fig. 3-24. Motion of a loop in a magnetic field

If we designate as $\Delta\Phi_1$ and $\Delta\Phi_2$ the fluxes cut by the sides *1* and *2* in time Δt, respectively, we may write the induced emf according to Eq. (3-26) as

$$e_1 = \Delta\Phi_1/\Delta t$$

and

$$e_2 = \Delta\Phi_2/\Delta t$$

At the end of the time Δt, the magnetic flux $\Delta\Phi_1$ will find itself inside and the magnetic flux $\Delta\Phi_2$ outside the moving loop (see Fig. 3-24). Because the direction of an emf is related to the applied magnetic flux by the corkscrew rule, e_2 will be in the positive direction and e_1 in the negative. Thus the net emf induced in the loop will be

$$\begin{aligned} e = e_2 - e_1 &= (\Delta\Phi_2 - \Delta\Phi_1)/\Delta t \\ &= -(\Delta\Phi_1 - \Delta\Phi_2)/\Delta t \\ &= -\Delta\Phi/\Delta t \end{aligned} \tag{3-27}$$

On replacing the elementary increments in the flux and time, $\Delta\Phi$ and Δt, by their infinitesimal changes, $d\Phi$ and dt, the emf induced in a loop at any arbitrary time can be written

$$e = -d\Phi/dt \tag{3-28}$$

Or, to state the above expression in words, *the emf of electromagnetic induction induced in a loop is equal to the rate of negative change of the magnetic flux linking the loop.* This is a second statement of the law of electromagnetic induction.

Experiments show that this change of flux can be caused either by the movement of a loop through a magnetic flux or by a change (positive or negative) in the magnetic flux linking the loop.

If the loop has w series-connected turns, the induced emf will be

$$e = -w\,(d\Phi/dt) \qquad (3\text{-}29)$$

Any closed contour (such as turns of a wire) links a definite amount of magnetic flux. This flux linkage is given by the product of the number of turns and the flux linked

$$\Psi = w\Phi \qquad (3\text{-}30)$$

So, the induced emf

$$e = -w\,(d\Phi/dt) = -d\Psi/dt \qquad (3\text{-}31)$$

is equal to the rate of negative change of flux linkage.

When a loop is moved in the direction indicated in Fig. 3-24, the magnetic flux linking the loop decreases, that is, its increment is negative, $\Delta\Phi < 0$, because $\Delta\Phi_2 > > \Delta\Phi_1$. As a consequence, the emf given by Eq. (3-28) is positive and has a clockwise sense. The current due to this emf has the same sense. The magnetic flux produced by this current is in the same direction as the decreasing magnetic flux, which can readily be proved by applying the corkscrew rule. To sum up, a decrease in the flux linking a loop gives rise to an emf and a current such that the resultant magnetic flux tends to oppose the decrease in the flux.

If we move the loop of Fig. 3-24 in the opposite direction, the flux linking the loop will build up ($\Delta\Phi > 0$), so the emf will, according to Eq. (3-28), be negative and have a counter-clockwise sense. The magnetic flux it produces will have the same sense. The magnetic flux produced by this current will be opposite to the growing flux due to the loop. To sum up, an increase in the flux due to a loop gives rise to an emf and a current such that the resultant magnetic flux opposes the increase in the flux due to the loop.

From the foregoing it follows that *the direction of an induced emf is always such that its current opposes the operation (or effect) that has caused it.* This law was formulated by H.F.E. Lenz of Russia in 1833.

If we increase the current traversing the solenoid of an electromagnet (see Fig. 3-25) or bring closer together a coil and a magnet, the magnetic flux linking the coil will build up, thereby giving rise to an emf and a current, i, in the coil. By Lenz's law, the direction of the magnetic flux

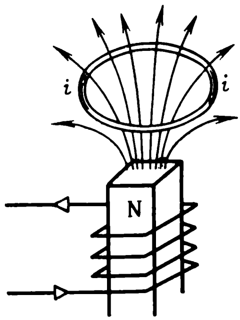


Fig. 3-25. Current induced in a ring

produced by the current i inside the coil is opposite to that of the flux due to the electromagnet, so the direction of the induced current can readily be determined by the corkscrew rule.

3-13. Operating Principle of an Electric Generator

When a wire is moved (see Fig. 3-26) in the direction of the velocity vector v in a plane perpendicular to the magnetic lines of force, an emf E is induced in the wire. This emf gives rise to a flow of current I around the closed circuit of resistance R. A current-carrying wire, when placed in a magnetic field, is acted upon by an electromagnetic force, $F = BIl$, whose direction, as given by the left-hand rule, is opposite to that of the velocity vector, so this is a retarding force.

Obviously, in order to move a wire one has to apply an external force equal in magnitude and opposite in direction to the retarding force. To state this differently, one needs a prime mover capable of delivering mechanical power $P_m = Fv$, or

$$P_m = Fv = BIlv = EI = P$$

Thus, the mechanical energy imparted to the wire to move it through the magnetic field is converted to electric energy, and a wire moved in a magnetic field by some mechanical force may be regarded as an elementary electric generator.

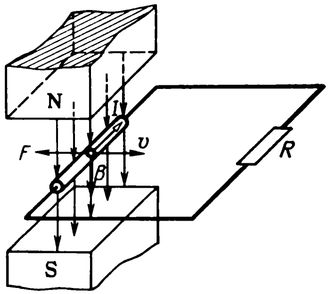


Fig. 3-26. Operating principle of an electric generator

Recalling Eq. (2-8), the emf of a generator may be written

$$E = V + V_0 = IR + Ir_0$$

So, the associated mechanical power

$$P_m = EI = I^2R + I^2r_0 = VI + P_0 = P_L + P_0$$

is equal to the electric power P which is the sum of the load power $P_L = VI$ and the power lost in the generator, $P_0 = I^2r_0$.

3-14. Operating Principle of an Electric Motor

If we place a wire of length l in a uniform field (see Fig. 3-27) so that it is at right angles to the magnetic lines of force, and pass through it a current I supplied by a source of voltage V, the electromagnetic force acting on the wire will, according to Eq. (3-1), be equal to

$$F = BIl$$

and its direction will be that given by the left-hand (wire) rule.

This force causes the wire to move with velocity v. In the process, the wire does a mechanical work, and an emf is induced in it. The direction of this emf, as given by the right-hand rule, is opposite to that of the current and its magnitude is

$$E = Bvl$$

If the wire has a resistance r_0, then, by Kirchhoff's voltage law, we may write

$$V - E = Ir_0$$

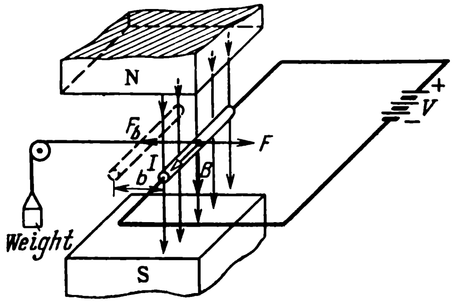


Fig. 3-27. Operating principle of an electric motor

or

$$V = E + Ir_0 \tag{3-32}$$

whence the current around the circuit is

$$I = (V - E)/r_0 \tag{3-33}$$

Multiplying both sides of Eq. (3-32) by the current I gives the electric power

$$VI = EI + I^2r_0 = BIlv + I^2r_0 = Fv + I^2r_0 \tag{3-34}$$

Here, the product I^2r_0 is the power lost as heat in the wires and Fv is the mechanical power.

Thus, the electric power generated as a wire is moved through a magnetic field is converted to mechanical power, and this process entails the induction of a counter-emf. The wire moved through a magnetic field may be regarded as a simple electric motor.

Example 3-8. Given: A wire 0.5 m long moved in a magnetic field of 1.2 T induction with a velocity of 20 m/s at right angles to the magnetic lines of force. The resistance of the wire is 0.1 ohm and the terminal (open-circuit) voltage is 15 V.

To find: (1) The power in the circuit; (2) the mechanical power developed by the wire; and (3) the power lost as heat.

Solution.

The counter-emf induced in the wire is

$$E = Blv = 1.2 \times 0.5 \times 20 = 12 \text{ V}$$

The current in the wire is

$$I = (V - E)/r_0 = (15 - 12)/0.1 = 30 \text{ A}$$

The power in the circuit is

$$P = VI = 15 \times 30 = 450 \text{ W}$$

The mechanical power is

$$P_m = EI = 12 \times 30 = 360 \text{ W}$$

The power lost as heat is

$$P_h = I^2 r_0 = 30^2 \times 0.1 = 90 \text{ W}$$

3-15. Eddy Currents

Figure 3-28 shows a metal disc mounted on a pivot and the traces of the poles of two electromagnets. The magnets produce magnetic fluxes, Φ_1 and Φ_2, which link the disc. Their magnetic induction vectors, B_1 and B_2, are shown in the same figure.

Any change in the current traversing the coil of one of the electromagnets brings about a change in the magnetic flux Φ_1, with the result that *eddy currents*, i_{e1}, are induced in the disc, similar to those induced in a ring-shaped coil (see Fig. 3-25). The direction of the eddy currents can be determined by the same rule as for a coil.

The interaction of the eddy currents i_{e1} with the magnetic flux Φ_2 gives rise to an mmf, F_1, which causes the disc to rotate.

Figure 3-29 shows the metal disc of a power meter and the trace of a pole of a permanent magnet. As the disc rotates it cuts the magnetic lines of force, and eddy currents i_e are induced in the disc. The direction of the emf induced in the disc and of the eddy currents which have the same sense is given by the right-hand (coil) rule.

The interaction of these currents with the field of the same permanent magnet produces an electromagnetic force and a retarding torque essential for the operation of a power meter.

Eddy currents can also be produced by changes in the magnetic fluxes linking the cores (Figs. 3-30a and 3-31a),

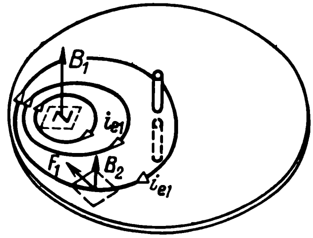


Fig. 3-28. Eddy currents caused in a disc by variations in a magnetic flux

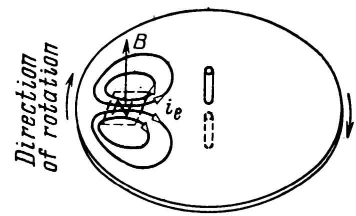


Fig. 3-29. Eddy currents induced by rotation of a disc in an unvarying magnetic field

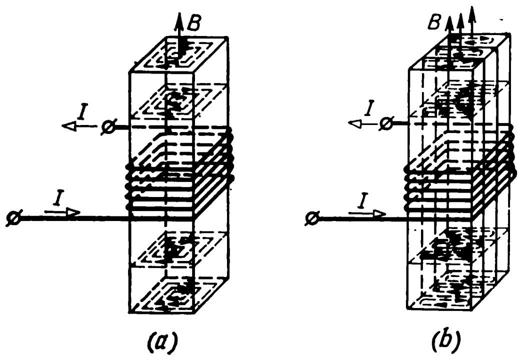


Fig. 3-30. Eddy currents in a steel core

enclosures and other parts of electric machines and apparatus. In such cases, eddy currents not only heat the metal they are flowing in, but also set up magnetic fields of their own, and these fields oppose the operation causing them. Heating by eddy currents constitutes a drain on the primary power source so this conversion of electricity to heat is quite appropriately called *eddy-current loss*. As a rule, specific eddy-current loss is of interest, that is, the loss of power per unit mass of iron, ordinarily expressed in watts per kilogram. Eddy currents can be utilized to advantage in electric furnaces and various heating appliances; but in electric machines and apparatus they entail an additional oss of power and bring down the efficiency.

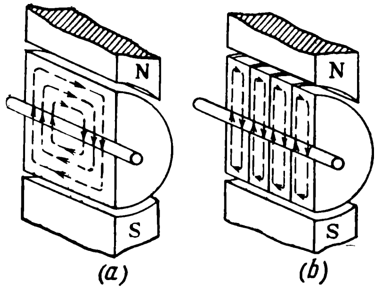


Fig. 3-31. Eddy currents in the armature of an electric machine
(a) solid armature; (b) laminated armature

One way to cut down eddy-current loss is to build cores from high-resistance steels, such as those containing 0.5 to 4.8 per cent silicon (electrical-sheet steel). Also, the cores of machines should be assembled from thin punchings or laminations (0.1 to 0.5 mm thick), insulated from one another (see Figs. 3-30*b* and 3-31*b*).

3-16. Inductance. EMF of Self-Induction

When a current is flowing around a circuit, each loop or turn of a coil links some amount of magnetic flux. This flux is called the *flux of self-induction* and symbolized as Φ_L. The sum of the fluxes of self-induction due to all turns of a loop or coil is referred to as the *flux linkage of self-induction*, Ψ_L. If the material has a constant permeability, the magnetic flux and flux linkage of self-induction are proportional to the flux-producing current.

The ratio of the flux linkage of self-induction to the flux-producing current in a loop or coil, with the permeability of the material held constant, is referred to as the *self-inductance* (or, simply, inductance) of the loop or coil

$$L = \Psi_L/I \tag{3-35}$$

As is seen, the self-inductance of a circuit relates its flux linkage of self-induction to the flux-producing current in the circuit.

The SI unit of self-inductance is the henry, H:

$$[L] = [\Psi_L/I] = \mathrm{Wb/A} = \mathrm{V\,s/A} = \Omega\mathrm{s} = \mathrm{H}$$

Fig. 3-32. Diagram symbol for an inductance

However, the henry is too large to be convenient for practical purposes. Instead, use is made of its submultiples, such as the millihenry, mH (1 mH = 1×10^{-3} H) and the microhenry, μH (1 μH = 1×10^{-6} H).

The diagram symbol for an inductance is shown in Fig. 3-32.

Let us determine the inductance of a ring-shaped coil. The flux linkage of this type of coil is given by Eq. (3-20)

$$\Psi_L = w\Phi = \mu_a (Iw^2/l) S$$

and its inductance is

$$L = \mu_a (w^2/l) S \tag{3-36}$$

As is seen, the inductance of a coil is a function of the coil size, number of turns, and core permeability.

Example 3.9. Given: A coil 30 cm (0.3 m) long, 5 cm (0.05 m) in diameter, carrying 2000 turns wound on a non-magnetic core ($\mu_a = \mu_0$).

To find: The inductance of the coil.

Solution.

The inductance of a coil is given by Eq. (3-36)

$$L = \mu_0 (w^2 S/l) = 125 \times 10^{-8} \times \frac{2^2 \times 10^6 \pi \times 5^2 \times 10^{-4}}{0.3 \times 4} = \text{(approx.) } 33 \text{ mH}$$

Any change in the current traversing a circuit (a loop) is accompanied by a change in the magnetic flux and flux linkage of self-induction, which obviously leads to the generation of an emf called the *emf of self-induction*. The process involved is known as self-induction.

The emf of self-induction is, according to Eq. (3-31), given by

$$e_L = -d\Psi_L/dt$$

or, replacing $d\Psi_L$ by $d(Li)$, we get

$$e_L = -d\Psi_L/dt = -d(Li)/dt = -L\, di/dt \tag{3-37}$$

Or, in words, *the emf of self-induction is proportional to the inductance of the circuit and the rate of change of the flux-producing current.*

The direction of the emf of self-induction is given by Lenz's law. When the flux-producing current is increased, that is, when $di/dt > 0$, the emf of self-induction, e_L, is negative and, as a consequence, opposes the flux-producing current. Conversely, when the flux-producing current is reduced, that is, when $di/dt < 0$, the emf of self-induction, e_L, is positive and, as a consequence, aids the flux-producing current.

Example 3-10. Given: A circuit with an inductance of 5 mH, traversed by a current whose rate of change is 600 A/s.

To find: The emf of self-induction.

Solution.

Because the current drops at the rate

$$-di/dt = 600 \text{ A/s}$$

the emf of self-induction is

$$e_L = -L\, di/dt = 5 \times 10^{-3} \times 600 = 3 \text{ V}$$

3-17. Energy of the Magnetic Field

When a d.c. source is connected to a circuit having a a resistance and an inductance, the circuit current gradually rises from zero to its final value given by

$$I = V/r$$

This rise is accompanied by the build-up of the surrounding magnetic field which stores some of the energy expended by the d.c. source. This energy manifests itself, for example, when the circuit is short-circuited, by maintaining the flow of current until all of the energy is expended to heat the circuit conductors. It also manifests itself through the interaction of the field with any current-carrying conductor that may be placed in that field.

In a coil, the rise of current defined above is accompanied by the generation of the emf of self-induction, $e_L =$

$= -L\, di/dt$. By Kirchhoff's voltage law, we may write

$$V + e_L = ir$$

whence

$$V = ir - e_L = ir + L\, di/dt \qquad (3\text{-}38)$$

As is seen, the terminal voltage of the circuit is the sum of two components, ir and $L\, di/dt$. The first component is given by Ohm's law. The second is equal in magnitude but opposite in direction to e_L; thus, it balances out the emf of self-induction arising in the circuit.

Multiplying both sides of Eq. (3-38) by the product $i\, dt$ gives

$$Vi\, dt = i^2 r\, dt + Li\, di$$

The left-hand side of the above equation gives the energy that the circuit receives in the time dt; the right-hand side of the same equation shows that some of the energy, $i^2 r\, dt$, goes to heat the circuit wire, while the remaining energy, $Li\, di$, is stored by the magnetic field round the circuit.

If we add together the energy increments occurring as the circuit current rises from zero to its final value I, we shall obtain the energy stored by the magnetic field round the circuit

$$W_m = \int_0^I Li\, di = LI^2/2 = \Psi I/2 \qquad (3\text{-}39)$$

3-18. Mutual Inductance

A change in the current flowing around a circuit (in a coil) will induce an emf in another circuit (or coil), if the two are placed in close physical proximity, or coupled, to each other. This process is known as *mutual induction.*

The current I_1 in the first coil (Fig. 3-33*a*) gives rise to a magnetic flux part of which, Φ_{12}, links the second coil, w_2, thereby producing the flux linkage of mutual induction, $\Psi_{12} = w_2\Phi_{12}$.

The magnetic flux Φ_{12} and, as a consequence, the associated flux linkage are proportional to the flux-producing

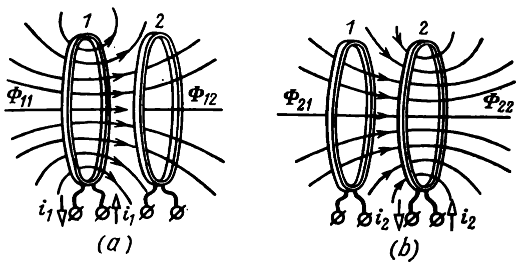


Fig. 3-33. Inductive coupling between two coils

current, I_1, that is

$$\Psi_{12} = M_{12}I_1$$

whence (3-40)

$$M_{12} = \Psi_{12}/I_1$$

The ratio of the flux linkage of one coil to the flux-producing current in the other is called *the mutual inductance between the two coils* (or circuits).

From a comparison of Eqs. (3-35) and (3-40) it follows that the unit of mutual inductance is the same as that of self-inductance — the henry, H.

The current I_2 in the second coil (Fig. 3-33*b*) produces a magnetic flux of its own, Φ_{21}, which links the turns of the first coil, w_1, thereby producing a flux linkage of mutual inductance, $\Psi_{21} = w_1\Phi_{21}$. As before, the flux linkage and mutual inductance may be written

$$\Psi_{21} = M_{21}I_2$$

whence (3-41)

$$M_{21} = \Psi_{21}/I_2$$

It is an easy matter to show that for two coupled coils or loops

$$M_{12} = M_{21} = M$$

always, so the subscripts "12" and "21" may safely be dropped.

The mutual inductance between coupled coils is a function of the number of turns, size, shape and relative position of the coils, and also the permeability of the medium.

A change in the current flowing in one coil brings about a change in the flux linkage of mutual induction, so, by the law of electromagnetic induction (see Sec. 3-12), an emf of mutual induction is induced in the other coil

$$e_2 = -d\Psi_{12}/dt = -M\,di_1/dt \tag{3-42}$$

Likewise, a change in the current flowing in the other coil brings about a change in the flux linkage of mutual induction, and emf of mutual induction is induced in the first coil

$$e_1 = -d\Psi_{21}/dt = -M\,di_2/dt \tag{3-43}$$

Or, to state this in words, *the emf of mutual induction is proportional to the mutual inductance between the coils and the rate of change of the flux-producing current.*

The mutual inductance between two coils is connected to the self-inductances of the same coils by a relation of the form

$$M = k\sqrt{L_1 L_2}$$

where k is the *coefficient of coupling* which defines the extent of inductive coupling between two coils.

The coefficient of coupling depends on the relative position of the coils. The shorter the distance between the coils, the greater the coefficient of coupling, and vice versa.

Mutual induction is utilized in various machines and apparatus, for example, to transfer energy from one circuit to another or to step up or down a voltage by means of a transformer.

Sometimes, mutual induction may be undesirable. For example, if a communication line runs parallel to a power transmission line, the emf of mutual induction induced in the communication circuit may seriously interfere with its operation.

3-19. Magnetohydrodynamic Generator

A magnetohydrodynamic (MHD) generator operates on the following principle.

Air, raised in temperature and enriched with oxygen, is admitted to a combustion chamber where a gaseous fuel is burned to produce a temperature of about 2500°C. The gaseous

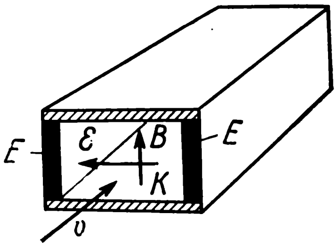


Fig. 3-34. Sketch of an MHD generator

plasma formed at this temperature has high electric conductivity. From the combustion chamber, the plasma is admitted at velocity v to an MHD duct, D (Fig. 3-34). Rectangular in cross-section, this duct is formed by two pairs of walls. One pair, E, are the metal electrodes of the MHD generator, and the other pair is made of a dielectric material. External electromagnets set up in the MHD duct a magnetic field of induction B, which is at right angles to the longitudinal axis of the duct.

When the plasma moves through the magnetic field down the duct (see Fig. 3-34), an electric field is set up in it, which is at right angles to the direction of motion of the plasma and the magnetic field. As a result, an emf is induced between the electrodes of the generator, whose magnitude is controlled by the induced electric field $\mathcal{E}$.

The power generated by an MHD generator is delivered to load over wires that connect the load to the generator electrodes. In operation under load, forces arise that oppose the motion of the plasma, so an increase in load inevitably slows down the plasma.

As we have seen, the thermal energy fed to an MHD generator is first converted to the energy of motion of the plasma, and this is then converted to electric energy.

The efficiency of a power plant incorporating an MHD generator may be as high as 60%, which is markedly greater than the efficiency of conventional thermal power plants (43%).

Chapter Four — Direct-Current Electric Machines

4-1. Functions

Electric machines are devices intended to convert mechal nical energy to electricity or electricity to mechanical energy. In the former case, they are called *electric generators*; in the latter, *electric motors*.

Direct-current (d.c.) generators are used to power electric motors, electrolysis cells, battery chargers, and the like. Direct-current (d.c.) motors actuate mechanisms which require large starting torques and speeds adjustable over a wide range, such as electric trains, mine hoists, and rolling mills. In automatic control systems, d.c. machines can be used as actuators, tachometers, signal converters, etc. In metal-working machine-tools, d.c. machines greatly simplify speed control.

4-2. Design of D.C. Machines

A d.c. machine operates on the principles set forth in Secs. 3-13 and 3-14. In sketch form, a two-pole (bipolar) d.c. machine is shown in Fig. 4-1. The machine consists of a steel frame, *1*, and a rotating armature, *2*. Bolted to the frame are poles, *3*. The poles (Fig. 4-2) carry a field (excitation) winding (at *4* in Fig. 4-1) having w_f turns and carrying a field (excitation) current I_f. The field winding gives rise to an mmf equal to its ampere-turns $I_f w_f$, which in turn sets up an excitation magnetic flux Φ_f, for which the closed path is formed by the poles, the air gap between the poles and armature, the armature and the frame (see Fig. 4-1).

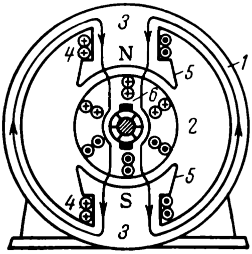


Fig. 4-1. Bipolar d.c. machine

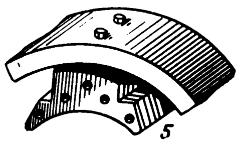


Fig. 4-2. Pole of an electric machine

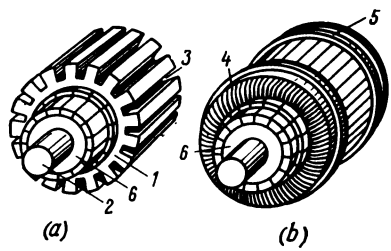


Fig. 4-3. Armature of an electric machine

The poles are built up of electrical-sheet steel laminations and have pole-pieces or pole-shoes, *5*, whose shape controls the distribution of the magnetic induction in the air gap B_δ.

The design of the armature is shown in Fig. 4-3. Its cylindrical body, *1*, is assembled from electrical-sheet steel punchings insulated from one another and press-fitted on a shaft or hub (at *2* in Fig. 4-3*a*). Slots, *3*, in the armature receive the wires (usually called "sides") of the armature winding (at *4* in Fig. 4-3*b*), which are connected in series-parallel. The armature winding is insulated from the slots and held in place by suitable wedges or binding wires, *5*.

The armature shaft also carries a commutator, *6*, electrically insulated from the shaft. The commutator (Fig. 4-4) is made up of copper commutator segments or bars, *1*, insulated from one another by micanite spacers which are

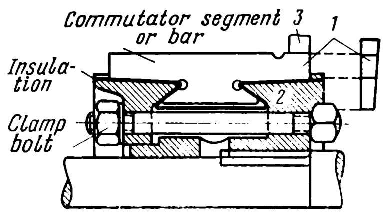


Fig. 4-4. Commutator

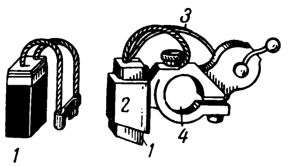


Fig. 4-5. Brushes and a brush-holder

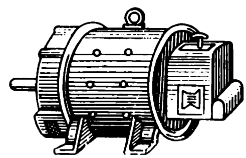


Fig. 4-6. External appearance of a d.c. machine

mounted on a sleeve, *2*, where they are held in place by bolts. The commutator bars or segments have risers, *3*, to which the wires of the armature winding are soldered in a definite order. Riding on the commutator surface are stationary carbon or graphite brushes (at *6* in Fig. 4-1) to which are connected leads from an external circuit. In this way, the external circuit is connected via the brushes and commutator to the rotating armature winding.

The arrangement of what is known as brushgear is shown in Fig. 4-5. The prismatic graphite (or graphitized) brushes, *1*, are held in the brush holders, *2*. Each brush holder is mounted on a stud passing through a hole, *4*, and supported by, but insulated from, the end shield of the machine. Attached to each brush are flexible copper pig-tails, *3*, which connect the brushes to the suitable armature terminals on a terminal board or panel. The terminal panel also carries terminals for the field winding (which may be of the shunt- or the series-wound type) and for the compole (or commutating-pole) winding (see Secs. 4-11 through 4-13). A general view of a d.c. machine is shown in Fig. 4-6.

4-3. Operating Principle of a D.C. Machine

A simplified circuit of a d.c. machine is shown in Fig. 4-7. The brushes are connected to a double-pole (double-throw) knife-blade switch, *1*, so that the armature can be connected to a load r or a supply line. The field winding, *2*, is connected to the supply line.

Let the armature connected to a load r be actuated by a prime mover, such as a heat engine. Then an emf E will be induced in the armature winding rotating in the field set up by the excitation (field) current I_f, and a current will be flowing through the load. The direction of the emf and current in the armature I_a, as given by the right-hand rule, is shown in Fig. 4-7. The direction of braking or retarding electromagnetic forces F_b that act on the current-carrying conductors in the magnetic field is likewise marked in the figure. These forces produce a braking (or retarding) torque on the shaft of the machine. The prime mover supplies a torque T which opposes the retarding torque. Thus, as has already been shown in Sec. 3-13, the machine operates as a generator converting mechanical energy to electricity.

By Ohm's law, the output current is

$$I = I_a = E/(r + r_a) \tag{4-1}$$

So,

$$E = Ir + Ir_a = V + Ir_a \tag{4-2}$$

or, stated in words, *the generator emf E exceeds the terminal voltage V by the voltage drop across the armature Ir_a.*

If we disconnect the shaft of the machine from its prime mover and place the switch in the upward position (see Fig. 4-7) the armature winding will be traversed by a current $I = I_a$ flowing in the direction opposite to that assumed before. The electromagnetic force produced by the interaction of this current with the magnetic field will likewise act in the opposite direction and produce a torque T driving the armature in the previous direction. Now the electric energy coming from the supply line will be converted to mechanical energy (see Sec. 3-14), so the machine will now operate as an electric motor.

The commutator and brushes switch the coils in the winding of the rotating armature in such a manner that

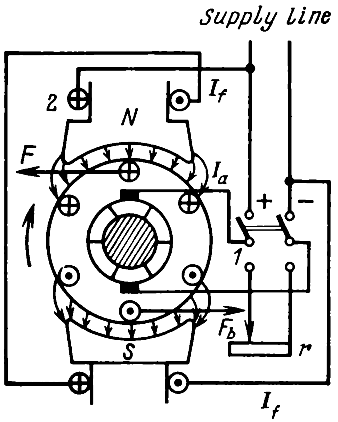


Fig. 4-7. Explaining the principle of a d.c. machine

each time the active conductors move from the north-pole belt into the south-pole belt the flow of current through them is reversed and the machine keeps rotating always in the same direction.

As with an electric generator, an emf is induced in the armature winding of the motor. However, this emf opposes the armature current, I_a, which can readily be proved by applying the right-hand rule. This is called the *counter* (or *back*) *emf.*

By Kirchhoff's voltage law,

$$V - E = I_a r_a$$

or

$$E = V - I_a r_a \tag{4-3}$$

so the armature current is

$$I_a = (V - E)/r_a \tag{4-4}$$

Or, in words, *when a machine is operating as a motor, its emf E is less than its terminal voltage V by the voltage drop across the armature winding, Ir_a.*

An electric motor can be reversed by reversing the direction of current flow in the armature or field winding. If the currents in the two windings are both reversed at the same time, the motor will keep rotating as before (see Fig. 4-7)

4-4. Construction of the Armature Winding

A simplified circuit of the armature winding is shown in Fig. 4-8.

Referring to the figure, there are two poles, N and S, between which is rotating an armature with six slots each of which receives two layers of armature coil sides wound with an insulated wire. For simplicity, the slots are not shown in the figure. As is shown in Fig. 4-8, the lead wire starting at commutator bar *1* runs over the near end of the armature to the top layer of the coil side in the first slot, away from the reader beyond the plane of the drawing, then, as shown by the dashed line, over the far end of the armature to the lower layer of the coil side in the fourth slot whence it is taken over the near end of the armature to the commutator bar marked *2*. The lead wire starting at commutator bar *2* is taken to the upper layer of the coil side in the second slot, and so on.

If we trace the run of the winding wire, we shall see that it closes on itself and consists of identical coils, the two ends of each coil being connected to adjacent commutator segments. This is known as the lap type of winding.

When the armature is rotating, an emf is induced in the coil sides enclosed in the slots, so they are called *active coil sides*. No emf is induced in the wires at the ends of the armature which are called *coil ends*. In sketch form a coil of the armature winding is shown in Fig. 4-9. The active sides of the upper layer are represented by a solid line, and that of the lower layer by a dashed line. To give them the desired shape, the coils are wound on suitable formers, then insulated, and dropped in the slots of the armature core.

Because each commutator segment receives two leads, namely the finish of the previous coil and the start of the next one, the number of segments K must be equal to that of coils in the armature winding. For the winding shown in Fig. 4-8, the armature has $Z = 6$ slots and as many coils. Once K is known, it is an easy matter to determine the number of active conductors that make up the armature winding

$$N = 2w_cK$$

where w_c is the number of turns per coil.

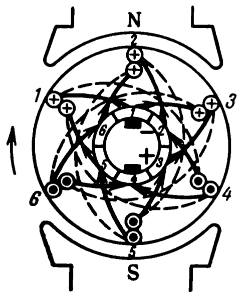


Fig. 4-8. Arrangement of the armature winding

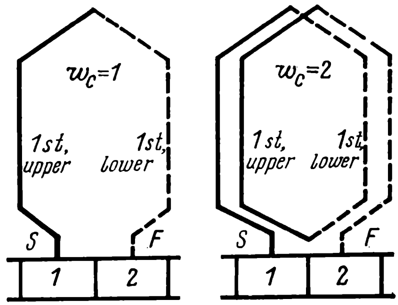


Fig. 4-9. Armature winding coil

The arrangement of armature windings can conveniently be studied by reference to the drawing in Fig. 4-10. This is the development of the cylindrical surface of the armature and its winding shown in Fig. 4-8. The directions of the emfs induced in the active conductors, as determined by the right-hand rule, are shown in Figs. 4-8 and 4-10. The sum of all emfs in a closed winding is zero. However, if we sum the emfs, starting, say, at the first commutator segment and moving in the direction of the emf, we shall notice that at the 4th segment the emf reverses its polarity. This is a junction point, or node, for two parallel branches of the armature winding, formed relative to the external circuit. As we keep moving down the winding against the emf, we shall run into a second node at commutator segment *1*, where the emf reverses its polarity again.

Thus, the winding consists of two parallel branches ($2a = 2$) with two nodes. The node at the 4th commutator segment is the point of the highest potential (+), and that at the first segment, the point of the lowest potential (—). The brushes are positioned to ride on these segments. The voltage between the two brushes at the time corresponding to the position of the armature shown in Figs. 4-8 and 4-10 is given by

$$v_1 = e_1 + e'_4 + e_2 + e'_5 + e_3 + e'_6$$
$$= e_4 + e'_1 + e_5 + e'_2 + e_6 + e'_3$$

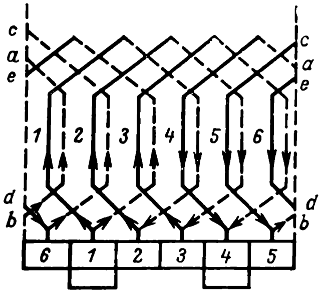


Fig. 4-10. Developed view of an armature winding

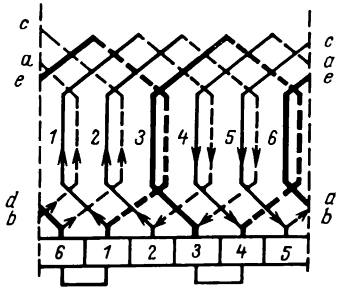


Fig. 4-11. Developed view of an armature winding

where the primed symbols stand for the emfs in the lower layers of the coils.

As the armature turns through 60°, the polarity of the brushes and the magnitude of v_1 remain as they were before, because the sixth slot takes up the position of the first, the first slot takes up the position of the second, and so on.

When the armature turns through an angle less than 60°, say, 30°, the winding takes up the position shown in Fig. 4-11, where the brushes, rather than the windings, are shifted to the left for convenience of study. In this position, two coils are short-circuited, and only two coils remain in each of the two parallel branches.

At this instant, the machine voltage is

$$v_2 = e_1 + e_4' + e_2 + e_5' = e_4 + e_1' + e_5 + e_2'$$

Thus, as the armature keeps rotating, its terminal voltage remains constant in polarity, but varies from v_1 to v_2 in magnitude. As the number of coils connected in each parallel branch is increased, the ripple in the output voltage is decreased. In present-day machines having a large number of coils, the ripple is so negligible that the output voltage may practically be taken unvarying.

The plane passing through the armature axis, that is, within an equal distance of the poles, and normal to their axis is called the *geometric neutral* (Fig. 4-12).

The short-circuited coils always move in the geometric neutral region where the induction B_δ is zero or negligible.

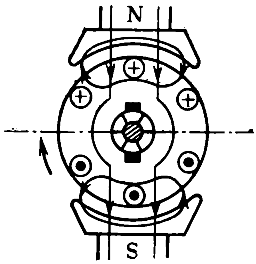


Fig. 4-12. Quadrature-axis armature reaction in a generator

Accordingly, the emf induced in the coils is equal to zero or negligible.

The armature windings of present-day machines are built essentially along the lines discussed above, irrespective of the number of poles, parallel branches, coils or commutator segments.

4-5. The EMF of the Armature Winding

As already noted, the emf of a machine is the sum of the emfs of the series-connected conductors in one parallel branch. These emfs are all different, because the magnetic induction is different at various points in the air gap around the circumference of the armature. However, the emf of a machine can be found in terms of the average emf of a conductor multiplied by the number of conductors in one parallel branch.

Let the magnetic flux due to one pole be Φ; the number of poles in the machine, $2p$; the axial length of the armature core, l; its diameter, d; and its surface area, S. Then the average magnetic induction on the armature surface will be

$$B_{av} = \Phi \times 2p/S = \Phi \times 2p/\pi dl \tag{4-5}$$

and the average emf of each conductor will be

$$E_{av} = B_{av}lv = \Phi \times 2p\pi\, dnl/\pi\, dl \times 60$$
$$= \Phi \times 2p\,(n/60)$$

where n is the rotational speed of the armature in revolutions per minute (rpm).

If the total number of conductors in the winding is N, and the number of parallel branches is $2a$, then each parallel branch will have $N/2a$ conductors connected in series. In the circumstances, the emf of a parallel branch and, as a consequence, that of the machine will be

$$E = E_{av}(N/2a) = 2p(n/60)(N/2a)\Phi$$
$$= (p/a)(n/60)N\Phi \tag{4-6}$$

or

$$E = c_E\Phi n \tag{4-7}$$

where $c_E = pN/(a \times 60)$ is the constant of the machine.

As is seen, *the emf of a machine is proportional to the magnetic flux and the rotational speed of the armature.*

4-6. The Electromagnetic Torque of a Machine

From Sec. 3-5 we know that a current-carrying conductor, when placed in a magnetic field, is acted upon by an electromagnetic force

$$F_{cond} = B_{av}Il = (\Phi \times 2p/\pi\, dl)\, l\,(I_a/2a)$$
$$= (p\Phi/\pi\, da)\, I_a$$

where B_{av} = average magnetic induction
d = armature diameter
l = armature length
$\Phi \times 2p$ = total flux of a multipole machine
$I = I_a/2a$ = current in a parallel branch, that is, the current in one conductor

The torque produced by each conductor in the armature winding is

$$T_{cond} = F_{cond}\, d/2 = (\Phi p/\pi\, da)(d/2)\, I_a$$
$$= (p\Phi/2\pi a)\, I_a$$

The total electromagnetic (or internal) torque of a machine having N conductors in the armature winding is given by

$$T = T_{cond} \times N = (p/2\pi a)\, N\Phi I_a = c_T\Phi I_a \tag{4-8}$$

where $c_T = (p/2\pi a)\, N$ is a constant.

When a machine is operating as a generator, then at load (with a current flowing in the armature winding), a retarding torque, T_r, is developed. When a machine is operating as a motor, a driving torque, T_d, is developed.

In either case, the electromagnetic torque T of a machine is balanced by a statical moment of resistance, M_s, and a dynamic moment, M_J, due to rotating masses, defined as

$$M_J = J\, d\omega/dt$$

where J is the moment of inertia, and the derivative $d\omega/dt$ is angular acceleration.

The moment of inertia is given by

$$J = m\rho^2$$

where m is the mass of the rotating body and ρ is the radius of inertia (or radius of gyration).

The equation of motion in this case has the form

$$T = -(M_s + M_J)$$

When the rotational speed $\omega = 2\pi n/60$ increases, the dynamic moment is positive; when the speed decreases, the dynamic moment is negative. If we take the components of the torque T, rather than the moments of resistance, then

$$T = M_s + M_J \qquad (4\text{-}9)$$

So long as $T > M_s$, a positive M_J is developed on the shaft, and the speed keeps building up. In the opposite case, the speed drops.

At a constant rotational speed (n = constant), the electromagnetic torque $T = c_T I_a \Phi$ is fully balanced by the statical moment which is the sum of the no-load torque T_0 necessary to overcome the rotational losses such as friction and the iron loss of the machine, and the load torque T_2, required to drive the associated mechanism, so

$$M_s = T_0 + T_2 \qquad (4\text{-}10)$$

4-7. Mechanical Power of a D.C. Machine

As already shown (see Secs. 3-13 and 3-14), when mechanical energy is converted to electrical energy or back, the mechanical power is EI. Let us prove that this also applies

to an electric machine. Suppose the force applied tangentially to the circumference of the armature is F and the linear velocity on the external surface of the armature is v. Then the total mechanical power will be

$$P_m = Fv$$

Substituting $F = 2T/d$ and $v = \omega d/2$ in the above equation gives

$$P_m = (2T/d)\,\omega\,(d/2) = T\omega \tag{4-11}$$

Because the torque of a machine is

$$T = (p/2\pi a)\,N\Phi I_a$$

the mechanical power is

$$\begin{aligned} P_m &= (p/2\pi a)\,N\Phi I_a\,(2\pi n/60) \\ &= \Phi N\,(p/a)\,(n/60)\,I_a \\ &= EI_a \end{aligned} \tag{4-12}$$

The product EI_a is often called *the electromagnetic power of a machine* and symbolized as P_e. When a machine is operating as a generator, EI_a exceeds VI_a, because $E > V$. When a machine is operating as a motor, VI_a exceeds EI_a, because $E < V$. The difference in power in either case is numerically equal to the power lost as heat in the armature winding, $I_a^2 r_a$.

4-8. Armature Reaction of a D.C. Machine

When a generator or a motor is operating at no-load, the armature current I_a is equal to zero or very small. In the circumstances, the magnetic flux Φ_f is produced solely by its mmf F_f and has its path completed through the armature along the poles. In Fig. 4-12, this flux is shown directed from top to bottom (from the N to the S pole).

In operation under load, the armature conductors carry a current I_a which in a generator (see Fig. 4-12) is in the same direction as the emf, given by the right-hand rule. Now the armature mmf F_a produces a magnetic flux of its own Φ_a, which has its path completed through the armature, the air gap, and the pole-pieces. Let us agree to call

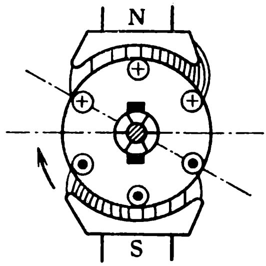


Fig. 4-13. Magnetic induction in the air gap in the presence of quadrature-axis armature reaction

the left-hand end of a pole-piece the leading edge (relative to the direction of rotation) and the right-hand end, the trailing edge. Then we shall see that the quadrature flux of the armature tends to demagnetize the leading edge and to magnetize the trailing edge of the pole-piece. The distribution of the magnetic induction in the air gap, B_δ, for this case is shown in Fig. 4-13. The resultant flux Φ of the machine is shifted in the direction of armature rotation, and so is the neutral plane of the machine, now called the *physical neutral plane* in contrast to the geometrical neutral plane.

Because of iron saturation (see Sec. 3-8), the demagnetizing action of the quadrature flux on the leading edge is greater than its magnetizing action on the trailing edge, so the total magnetic flux of the machine is reduced ($\Phi < \Phi_f$). The effect of the armature mmf on the magnetic flux of the machine in operation under load is called *armature reaction*.

When a machine is operating as a motor, the armature current (see Fig. 4-12) opposes the emf, the armature is rotating in the opposite direction as compared with operation as a generator, and the quadrature-axis armature reaction shifts the resultant flux Φ and the physical neutral plane against the direction of armature rotation.

To counteract armature reaction, it is customary to provide a *compensating winding* on the main poles. The conductors of the compensating winding are placed in slots on the surface of the pole-pieces. This winding is connected in series with the armature winding so that the current flowing in it opposes the current in the armature conductors.

4-9. Commutation

It has been explained in Sec. 4-4 that for the emf E to be constant in direction and nearly so in magnitude, the armature winding must be constructed so that it has $2a = 2$, 4, 6 and so on parallel branches or paths, and the coils are continually switched from one to the other. The switching action is effected mechanically by the commutator and brushes, and the action itself together with the physical events accompanying it are called *commutation*. The coil being commutated is short-circuited by a brush and finds itself near the geometric neutral. The emf induced in the coil by the external field of the poles is zero very nearly. The time T during which the switching action takes place and a coil remains short-circuited is a few thousandths of a second; it is called the *commutation period.*

Let us examine commutation in simplified terms.

The coil of interest (see Fig. 4-10) is found in slots *6* and *3* and is shown separately in Fig. 4-14. Suppose that the winding is rotating very slowly ($T \to \infty$), the brush width is equal to that of a commutator segment, and account is solely taken of the contact resistance between the brush and the commutator segment, $r_s = R$. Then the armature current I_a flows from the brush into commutator segment *1* and is divided into two equal parts, $i = 0.5\ I_a$. Figure 4-14 shows that the coil connected in the parallel branch going to the left carries a current $i_c = 0.5\ I_a$, which flows in the counter-clockwise direction.

During the next instant of time, when commutator segment *6* touches the brush, the current I_a will be divided in a different manner. If, for example, after a time $t = 0.1\ T$, the brush touches commutator segment *6* with one-tenth of its contact surface, the current flowing through that segment will be $i_6 = 0.1\ I_a$ and that through segment *1*, $i_1 = 0.9\ I_a$. As before, the currents in the parallel branches must be constant and equal to $0.5\ I_a$, if I_a is constant. Then, flowing as it did before, the coil current will be $i_c = i_1 - 0.5\ I_a = (0.9 - 0.5)\ I_a = 0.4\ I_a$ in one parallel branch, and $i_6 - i_c = (0.1 + 0.4)\ I_a = 0.5\ I_a$ in the other parallel branch. The coils being switched at the time under consideration are shown in Fig. 4-15. If we assume the time

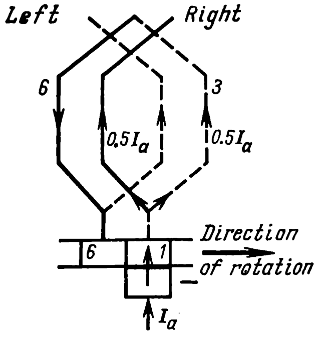


Fig. 4-14. Start of a commutation cycle ($t = 0$)

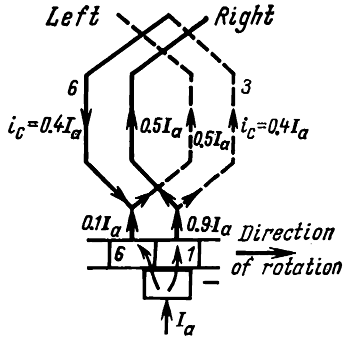


Fig. 4-15. Commutation at $t = 0.1T$

$t = 0.5\ T$, we shall learn that $i_c = 0$, and the currents in the parallel branches are $0.5\ I_a$ as before. With time, the coil current builds up in the opposite direction until at $t = T$ it becomes equal to $i_c = -0.5\ I_a$. At that time, the coil will be switched into the other parallel branch, and a cycle of commutation will be complete (Fig. 4-16). A plot of i_c as a function of time t yields a straight line (Fig. 4-17) which represents what is known as *linear commutation*; in designing a machine one always seeks to ensure linear commutation.

However, in practice the commutation time T is small, the coil current i_c changes at a high rate, and an emf of self-induction e_s is induced in it. Because $i_c = f\ (t)$ yields a straight line, which means that $di/dt = \tan \alpha =$ constant, then $e_s = -L_c di_c/dt$ is a constant, and an additional current arises in the coil, given by

$$i_s = e_s/r_s = e_s/\ (r_6 + r_1)$$

where r_6 and r_1 are the contact resistances between the brushes and the respective commutator segments. Once we know r_s for different values of t, we can readily determine i_s. For $t = T/2$, say, $r_s = r_6 + r_1 = 2R + 2R = 4R$. For $t = 0$ and $t = T$, r_s is infinity. A plot of i_s as a function of t is shown in Fig. 4-17b. The total current $i_c + i_s$ is shown in Fig. 4-17a by a dashed curve from which one can see that

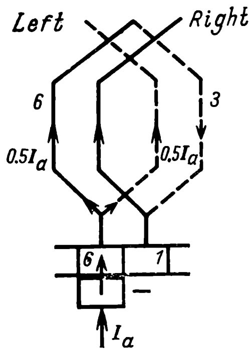


Fig. 4-16. End of a commutation cycle ($t = T$)

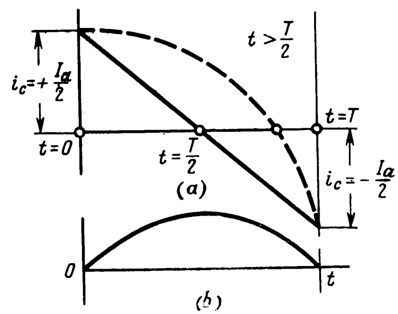


Fig. 4-17. Variation in the current of a commutated coil under natural conditions

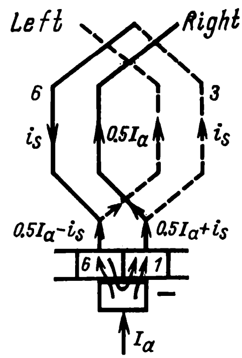


Fig. 4-18. Delayed commutation

in the presence of e_s the total coil current makes a zero-crossing at a later time than it should ($t > T/2$). This is known as *delayed commutation.*

In the case of delayed commutation, the current density in the trailing edge of a brush increases considerably (see Fig. 4-18), and this causes excessive heating and wear of the brush. However, the principal danger lies in the fact that delayed commutation is accompanied by sparking or even arcing between the commutator and the trailing edge of the brush. This occurs because when the coil circuit is opened, the electromagnetic energy, $L_c = i_s^2/2$, stored in it is dis-

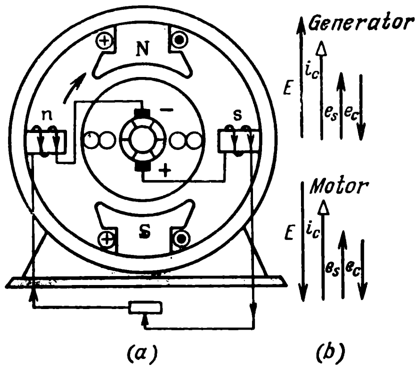


Fig. 4-19. Commutating poles

sipated by the electric arc that strikes at the trailing edge.

Should armature reaction build up the magnetic induction B_δ at the pole edge by 30 to 50%, the potential difference between the commutator segments of the coil moving in the region of elevated induction may be in excess of 25 to 35 V. This voltage is capable of sustaining individual arcs which, on merging in the ionized space around the commutator, may form a strong arc between the brush yokes of different polarity. This is known as the *commutator ring fire*; it may cause a serious damage or even a breakdown.

Commutation can be improved by several measures which have as their objective to minimize the additional current i_s. An effective one is to provide commutating poles, also known as interpoles.

Figure 4-19 shows a bipolar generator which, in addition to the main poles labelled N and S, has two commutating poles n and s arranged to lie on the geometric neutral and alternate, as shown, in the direction of armature rotation. The excitation (field) winding of the commutating poles is connected in series with the armature winding. The figure shows commutation taking place in two coils marked by filled circles. The emf E of the machine and the decreasing coil current i_c are directed aiding, as is the emf of self-induction, e_s, so it maintains the decreasing current (Fig. 4-19*b*). When the commutating poles have the polarity

shown in Fig. 4-19a, an additional commutating emf, e_c, is induced in the active coil sides. Its direction is such that it opposes the action of E and, as a consequence, that of e_s. If e_c is numerically equal to e_s, the additional coil current will be

$$i = (e_s - e_c)/r_s = 0$$

In a motor, the main and commutating poles will alternate in the direction of rotation in the N-n-S-s sequence.

Because e_s is proportional to I_a, the commutating poles are made unsaturable so that e_s can be balanced out at any load. Then $e_c \backsim \Phi_{cp} \backsim I_a$.

The severity of commutation can be estimated "by eye" as follows:

— severity degree 1: there is no sparking, blackening on the commutator, or carbon deposit on the brushes;

— severity degree $1^1/_4$: there is moderate sparking at a small part of the brush, no blackening on the commutator, or carbon deposit on the brushes;

— severity degree $1^1/_2$: moderate sparking at a greater area of the brushes, there are traces of blackening on the commutator (readily removable with a rag moistened with petrol) and those of carbon deposit on the brushes.

4-10. Ratings and Characteristics of Electric Machines

The frame of each electric machine carries in a conspicous place a name (or rating) plate which gives the duty for which the machine has been designed, its rated power, voltage, current, rotational speed, efficiency, etc., collectively called machine ratings.

A rating, or a rated value, indicates the limit(s) established by the maker (or designer) for application of a machine under specified conditions. In this text, all ratings, or rated values, will have the subscript "n" (because rated values are often called "nameplate or nominal values"). For example, the rated power will be designated as P_n, the rated voltage V_n, the rated current I_n, and so on.

Knowledge of machine ratings is essential for the intelligent use of the machine during its established service life

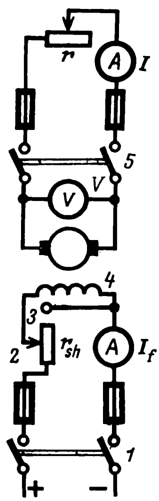


Fig. 4-20. Circuit of a separately excited generator

(about 10 years for stationary machines under normal ambient conditions).

When using a machine, it is important to know how some of its variables are related to other variables. Such relationships are usually referred to as characteristics. For example, the dependence of the terminal voltage of a generator on its load (circuit current) at constant rpm (n = constant) and constant field current (I_f = constant) is known as the *load* (or *external*) *characteristic*, $V = f(I)$, of the generator (Fig. 4-22). The dependence of the rotational speed of a motor n on the mechanical load on its shaft (the power P_2 or torque T) at constant armature voltage and constant field current is known as the *mechanical characteristic* of the motor (see Fig. 10-20).

Data for such characteristics are obtained by experiments or from past experience, and the values used in the respective equations are current in amperes, voltage in volts, etc. For practical purposes, however, it is more convenient to use percentage plots. To this end, experimental data are converted to percentages. Then $I\% = I/I_n \times 100$; $V\% = V/V_n \times 100$; $P\% = P/P_n \times 100$; $n\% = n/n_n \times 100$, etc.,

where I, V, P and n are found by experiments or from experience, while I_n, V_n, P_n and n_n are the nameplate (rated) values. The plots of Figs. 4-22, 10-19 and 10-20 have been constructed in precisely this manner.

Often, plots are constructed in per-unit quantities labelled I^*, V^*, P^*, etc. The per-unit value of a quantity is obtained by dividing its per cent value by 100, for example: $I^* = I\%/100$; $V^* = V\%/100$, etc. The per-unit values are convenient in circuit calculations.

4-11. The Separately Excited Generator

In a separately excited generator (Fig. 4-20) the field winding, *4*, is connected to an external source via a knife-blade switch, *1*, an ammeter and a field rheostat, *2*, which has an idle contact, *3*, connected to the field winding. If we reduce I_f by moving the contact arm of the rheostat upwards, it will reach the idle contact and short-circuit the field winding. If there were no idle contact, the field winding would be open-circuited, and an arc would strike at the circuit break, dissipating electromagnetic energy.

Since the inductance of the field winding is usually high, the emf of self-induction that would maintain a considerable voltage at the ends of the open-circuited field winding would also be high. This might cause insulation breakdown and grave danger to attending personnel. The arc would melt and oxidize the contacts.

The armature terminals receive connections from the load, r, a voltmeter, V, and an ammeter, A. The drive motor actuating the armature is not shown.

The *no-load characteristic* $E_0 = f(I_f)$ at n = constant and $I_a = 0$ is, in fact, the $\Phi_f = f(I_f)$ curve, plotted on a different scale, because E is proportional to Φ. It is used for a check on the design of the magnetic circuit and various graphic construction and is called *the magnetic characteristic of a machine.*

It is plotted as follows. The armature is rotated at a constant speed $n = n_n$, with the switches *1* and *5* open. Then switch *1* is closed, and the resistance of r_{sh} is brought down so as to raise the generator voltage to $V_0 = 1.1\ V_n$ to $1.2\ V_n$. On having written the values of I_f and V_0, one then re-

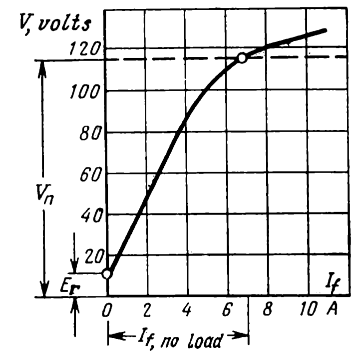


Fig. 4-21. No-load characteristic of a generator

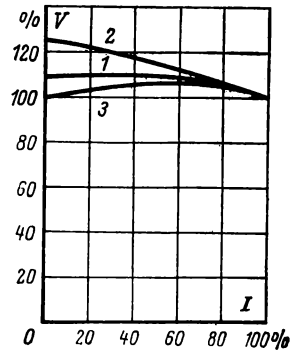


Fig. 4-22. External (load) characteristic of a generator

duces the excitation current consecutively and takes five or six readings of I_f and V_0. The last reading is taken at $I_f = 0$. The residual induction thus left induces what is known as the residual emf, E_r, usually equal to 2-2.5% V_n. The point representing V_n usually lies on the kink of the characteristic curve, and its abscissa is equal to the no-load excitation current at rated voltage, designated $I_{f,\ \text{no-load}}$ (Fig. 4-21).

The *load* (or external) *characteristic* of a generator, $V = f(I)$ at $I_f = \text{constant}$ and $n = \text{constant}$ relates variations in the generator voltage to variations in its load; it is shown in Fig. 4-22 as curve *1*.

To plot the load characteristic, the armature is driven at a constant speed $n = n_n$, and the machine is excited at no-load to obtain $V = 1.1\ V_n$ to $1.2\ V_n$. The switch *5* is then closed (see Fig. 4-20) and the load resistance r is gradually reduced while adjusting the exciting current so that, at $I = I_n$, the rated voltage V_n is obtained. This gives the first point of the characteristic for V_n and I_n (Fig. 4-22). Then, while holding I_f and n constant, one consecutively increases the load resistance r and takes five or six readings of V and I, while reducing the load on the generator to zero.

As the load is reduced, a decrease takes place in the armature current I, the voltage drop across the armature Ir_a, and the demagnetizing action of armature reaction (see

Sec. 4-8). The machine flux Φ rises, and so does the emf E. Because $V = E - Ir$, the terminal voltage V increases.

The quantity

$$\Delta V\% = \frac{V_0 - V_n}{V_n} \times 100\% \qquad (4\text{-}13)$$

is called *the voltage regulation of a machine.* For separately excited generators, it is usually 5 to 10 per cent.

To maintain voltage constant against variations in load, one has to adjust excitation current. This can be done manually or automatically. Separately excited generators are used in circuits intended to control the speed of motors between wide limits (see Sec. 4-16), in automatic control systems, and also where $V_a \neq V_f$.

4-12. The Shunt-Wound Generator

In a shunt-wound generator, the field winding is connected to the armature terminals in parallel with the load circuit (Fig. 4-23). In this case, the current carried by the armature is

$$I_a = I + I_f$$

where the excitation (field) current varies from one to seven per cent of the rated current.

To excite a shunt-wound generator, it is essential that the magnetic flux produced by the exciting current be in the same direction as the flux produced by the residual induction. It is only then that the current in the field winding due to the residual emf, E_r, will magnetize the machine, the magnetic flux of the generator will rise, and so will its emf. The latter will cause a further increase in I_f and, as a consequence, in Φ. This process of self-excitation (or "build-up" as it is usually called) goes on until the emf becomes equal to the voltage drop across the field winding

$$E = I_f r_f$$

If the generator fails to build up, it is necessary to reverse the exciting current I_f.

The no-load characteristic of a shunt-wound generator has the same form as that of a separately excited generator and can be used for similar purposes.

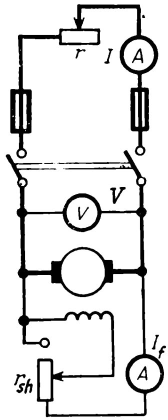


Fig. 4-23. Shunt-wound generator

The load (external) characteristic of a shunt-wound generator $V = f(I)$, at n and r_f being constant, is plotted in the same manner as for a separately excited generator and has the same shape (curve *2* in Fig. 4-22). However, the voltage regulation $\Delta V\%$ in this case is greater, being as high as 30%. The point is that the field winding of a shunt-wound generator is connected to the armature terminals. When load is shed, the voltage rises, and so does the exciting current $I_f = V/r_f$, so the magnetic flux and emf E of the machine increase by a greater amount than they do in a separately excited machine. Shunt-wound generators are used on a large scale, because they do not require an additional source for excitation.

4-13. The Compound-Wound Generator

The circuit of a compound-wound generator is shown in Fig. 4-24. It has two field windings, a shunt winding and a series winding, the latter being connected in series with the armature winding. With this connection of the series field winding, its mmf, $F_{m,s}$, can be added to or subtracted from that of the shunt winding, $F_{m,sh}$:

$$F_m = F_{m,sh} \pm F_{m,s}$$

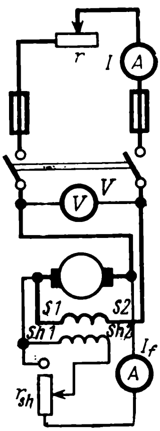


Fig. 4-24. Compound-wound generator

In practice, it is more customary to connect the two field windings in an aiding fashion, that is, so that their mmfs are added together. Then, the generator automatically builds up as load increases. The demagnetizing effect of armature reaction and the voltage drop across the armature winding Ir_a cancel out each other, and the voltage of such a generator remains practically constant despite variations in load (see curve *3* in Fig. 4-22). Compound-wound generators are used in applications where it is essential to have a nearly constant voltage under conditions of a frequently varying load.

When the two field windings are connected in opposition, their mmfs are subtracted from each other, and the load characteristic is then a steeply falling curve. Compound-wound generators in which the series field winding is connected in opposition are used in applications where a short-circuit is a normal condition, such as in welding.

4-14. Direct-Current Motors

Because one and the same electric machine can operate as a generator or a motor, the voltage drop across its armature winding at rated current will naturally be the same. This voltage drop is $I_n r_a = 4$ to 10% of the rated voltage. As a consequence, if we apply a voltage $V = V_n$ to the termi-

nals of a stationary armature, its current will be 25 to 10 times the rated current. Neither the armature winding nor the brushes or commutator are designed to withstand such a heavy current. That is why, the voltage across the armature, in the case of starting at rated current, must be brought down by as much as 90 to 96% of its rated value by connecting a rheostat r_s in series with the armature. This rheostat intended to limit and adjust the starting current, is termed a *starting rheostat*. The starting current may be as high as 200 to 250% of the rated current. Hence, the starting current is

$$I_s = V_n/(r_a + r_s) \tag{4-14}$$

whence it is an easy matter to determine the value of r_s.

When the armature is rotating, a counter-emf is induced in its winding and, as a consequence, the armature current

$$I_a = (V - E)/(r_a + r_s)$$

goes down as the motor picks up speed. The rheostat is no longer needed, and it is gradually brought out of circuit in steps, until its resistance is zero. At that instant, the armature current becomes

$$I_a = (V - E)/r_a \tag{4.14a}$$

If a motor is running idle (at no-load), the counter-emf may be as high as 99% of the rated voltage at a no-load current of $I_a = I_{no\text{-}load} =$ (approx.) 5 to 10% of the rated current. In operation under load, if $I_a = I_n$, the counter-emf is 90 to 96% of the rated value.

The rpm of a motor can be found by Eq. (4-15)

$$n = (1/c_E)\,(E/\Phi) \tag{4-15}$$

As is seen, *the rpm of a motor is proportional to the emf induced in the winding and inversely proportional to the magnetic flux*. Because $E = V - I_a r_a$, so

$$n = (1/c_E)\,\frac{V - I_a r_a}{\Phi} \tag{4-16}$$

As load on the shaft, that is, output power P_2, is increased, the terminal power, that is, input power P_1 of a motor increases. In the circumstances, if $V =$ constant, the armature current must rise. From Eq. (4-16) it is seen that the rpm of the motor must decrease. On the other hand, an in-

crease in I_a entails an increase in the demagnetizing action of armature reaction, so the rpm must rise. For the consistent operation of a motor, it is built so that the effect of the voltage drop across its armature, $I_a r_a$, and of a stabilizing winding (see Sec. 4-17) always prevails over that of armature reaction, and the rpm usually drops somewhat as load increases.

Recalling Eqs. (4-8) and (4-12), the torque and electromagnetic power developed by a motor, including friction loss, can be written

$$T = c_T I_a \Phi$$

and

$$P_{em} = EI_a = c_E I_a \Phi n$$

As the retarding torque T_2 increases, the driving torque T on the shaft is automatically raised owing to the increase in I_a until the two torques become equal, $T = T_2$, at some particular value of n. Thus, for each value of load, there is a definite rotational speed.

Like generators, motors are classed into shunt-wound, separately excited, series-wound and compound-wound types. While in generators one is above all interested in electric performance, in motors we are mainly interested in mechanical characteristics, such as $n = f(I)$, $n = f(I_f)$, $T = f(I)$, etc.

4-15. The Shunt-Wound Motor

A shunt-wound motor is best to drive mechanisms that require a nearly constant rotational speed and economical speed control. A simplified circuit of such a motor is shown in Fig. 4-25.

In the diagram, the letters L, F and A label the terminals of a starting rheostat, respectively connected to the supply line, the field winding, and the armature. The filled circles mark the contacts effectively used, and the white spaces between them represent the sections (or steps) in the resistance element of the starting rheostat. When the motor is running, a jumper, *3*, permanently connects the terminal L to a shunt rheostat intended to adjust the exciting (or field) current I_f. Before closing the knife-blade switch, it is important to see

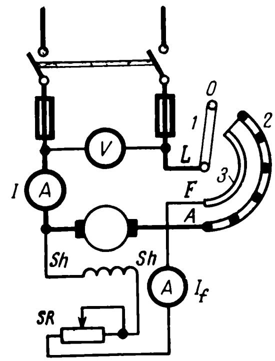


Fig. 4-25. Shunt-wound motor

that the movable contact *1* of the starting rheostat *2* resides on the idle contact 0. The movable contact of the shunt rheostat, designated *SR* in the diagram, must be in the leftmost position where its resistance is a minimum.

When the switch is closed and the starting rheostat is moved to the first step, the motor current I is divided into two parts, one flowing through the armature, I_a, and the other through the field winding, I_f.

Thus, the total, or supply, current is the sum of two terms

$$I = I_a + I_f \tag{4-17}$$

where I_f is 1 to 7% of the rated current.

Depending on the resistance of the starting rheostat, the initial current inrush, or the starting current I_s, may be as high as 200 to 250% of the rated current. The starting torque causes the armature to start rotating. As the motor picks up speed, the armature current goes down, and the starting rheostat may be moved to the second step. This brings about a sudden increase in the armature current and, as a consequence, an increase in torque and a further increase in speed, after which the armature current tends to go down again. Now the starting rheostat is moved to the third step, and so on. The starting phase is complete when the full rated voltage V_n is impressed on the armature winding. The starting rheostat has a resistance element usually designed for short-time duty at starting, and the movable contact ought not to be left at intermediate steps for a long time.

The faster the rate at which the armature counter-emf

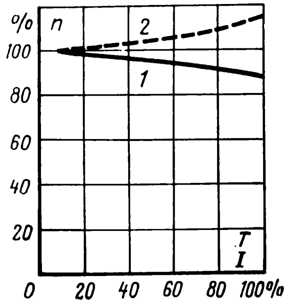


Fig. 4-26. Speed characteristics of a shunt-wound motor

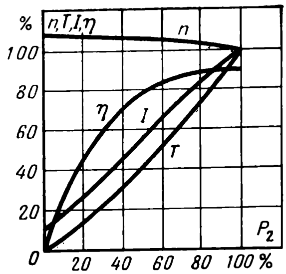


Fig. 4-27. Operating characteristics of a shunt-wound motor

rises, the faster the decrease in current and the smaller the heat build-up in the armature. This is the reason why motors are always started at a maximum exciting (field) current, by shorting out the resistance of the starting rheostat (Fig. 4-25). In the circumstances, the magnetic flux and the counter-emf of the machine are a maximum. Also, a motor must develop a high torque at starting, and this can likewise take place only at a maximum magnetic flux [see Eq. (4-8)].

To stop a motor, the starting rheostat is first moved to the idle (zero) step, and the switch is then opened. In this way, the switch contacts are protected against burning and the motor is set up in readiness for the next start.

Figure 4-26 shows the *speed characteristic*, $n = f(I)$, of a motor at V and I_f constant (curve *1*). In the absence of mechanical load, the no-load current is less than 10% of the rated current, and the speed is a maximum

$$n_{no\text{-}load} = (1/c_E)\frac{V - I_{no\text{-}load}r_a}{\Phi}$$
$$= \text{(approx.)}\ (1/c_E)\,(V/\Phi) \qquad (4\text{-}18)$$

because $I_{no\text{-}load}r_a = \text{(approx.)}\ 0$.

As load or, which is the same, load torque is increased, the speed of the motor remains practically unchanged (it falls off insignificantly). This happens because the armature resistance r_a is low, and even a small decrease in the counter-emf brings about a sudden rise in armature current ac-

cording to Eq. (4-14a), so the torque is automatically restored to its previous value. This form of response is known as a *flat* characteristic.

If the exciting (field) current is held constant, the magnetic flux may also be deemed constant, because the effect of armature reaction is negligible. Then the torque of the motor

$$T = c_T I\Phi = \text{(approx.)}\ (c_T\Phi)\, I \qquad (4\text{-}19)$$

is approximately proportional to the current I. So, if we lay off T as abscissa (Fig. 4-26), we shall obtain the mechanical characteristic of the motor

$$n = f\,(T)$$

at V and I_f held constant.

It is convenient to use the *operating characteristics* (Fig. 4-27) included in the data sheets and specifications of motors. This term applies to relationships of the form

$$n,\ T,\ I,\ \eta = f\,(P_2)$$

for V_n and I_f held constant, where η is the efficiency of the motor and P_2 is its useful power available at the shaft.

The power available at the shaft is

$$P_2 = T \times 2\pi n/60$$

and the torque is

$$T = P_2 \times 60/2\pi n \qquad (4\text{-}20)$$

If n were constant, the $T = f\,(P_2)$ characteristic would be a straight line passing through the origin of coordinates. However, n falls off as P_2 rises, so T is not proportional to P_2. If V is held constant, I is proportional to the input power, $P_1 = VI$. Because the power loss, $P_1 = P_2$, in the motor is small, I is approximately proportional to P_2.

The speed of a shunt-wound motor is usually regulated by adjusting its exciting (field) current. This method provides for accurate and continuous control in the ratio 1 : 1.5, or even 1 : 8, in special-purpose motors. This is done as follows. The torque of the motor $T = c_T I\Phi$, at Φ=constant, is proportional to I, and the latter is given by

$$I = (V - E)/r_a$$

Because r_a is small, the voltage drop across the armature circuit, $I_a r_a$, is also small. So, when V and r_a are constant, the armature current may rise appreciably even if the counter-emf decreases by a small amount.

For example, if $r_a = 0.5$ ohm, $V = 220$ V and $I_a = 10$ A, the counter-emf will be $E = V - I_a r_a = 220 - 10 \times \times 0.5 = 215$ V. If the counter-emf is brought down by a mere 10 V (about 5%), so that $E' = 205$ V, the armature current will be $I'_a = (220 - 205)/0.5 = 30$ A, that is, there will be a three-fold increase in its value.

Thus, if we bring down the exciting current by, say, 5% while holding both the load ($T_d = T_r$) and rpm constant, the magnetic flux and the counter-emf will decrease by the same amount. This will bring about a sudden increase in armature current and torque, with the excess torque going to accelerate the armature. However, as the armature picks up speed, the counter-emf rises again, and the armature current will fall off to a value at which $T = c_T \Phi I_a$ takes on its former value. In this way, at $T_d = T_r$, the motor comes up to a new constant speed which is higher than the previous one.

With this form of control, the field rheostat dissipates very little power (the power lost is $P_{rh} = I_f^2 r_f$), because I_f does not exceed 1 to 7% of I_n.

This form of control allows the motor speed to be raised above the rated value.

If, with the load on the motor shaft held constant, we place a series resistor r_s in the armature winding circuit, the armature current will at first go down, and so will the torque. As a consequence, the load torque will exceed the motor torque, and the speed will be forced down. On the other hand, the reduction in speed and counter-emf will cause the armature current to rise, and this will be accompanied by a rise in the motor torque. This chain of events will continue until the two torques balance each other — at that instant the speed ceases falling off, and the motor will keep running at a constant, although reduced speed. This form of control is bad because a substantial proportion of energy is dissipated by the series resistor.

4-16. The Separately Excited Motor

A separately excited motor is very close in its characteristics to a shunt-wound motor. Since, however, the field and the armature windings are energized by separate sources, rather than by a common supply line, it is possible to control motor speed accurately over a wide range of values by varying the voltage across the armature terminals and by use of what is known as across-the-line, direct-on-line, or full-voltage starting, that is, one in which no resistor or rheostat is interposed in series with the armature winding circuit.

A likely arrangement for operation of a separately excited motor is shown in Fig. 4-28. As is seen, the arrangement is a combination of an armature-controlled motor and a generator (usually referred to as Ward-Leonard system). The armature of the motor, *2*, is connected directly, that is, without a starting rheostat, to the armature of the generator, *1*, that energizes the motor. The generator and the motor are separately excited by an exciter, *6*. As a rule, the generator and the exciter are driven by an induction motor *7*.

The rpm of the armature-controlled motor can be varied by adjusting its exciting (field) current with a rheostat, *3*, and also by varying the voltage, V, of the generator, likewise through the adjustment of its exciting current with another rheostat, *5*.

When the exciting current of the main generator is reversed with a switch, *4*, the polarity of its brushes is also reversed, and so is the direction in which the controlled motor is running.

The Ward-Leonard system is used in mine hoists, marine propulsion units, and the drive of some metal-cutting machines. The mechanical characteristics of the system (Fig. 4-29) are very close to those shown in Fig. 4-26. Curve *1* in Figs. 4-26 and 4-29 is the natural characteristic that would be obtained if the system operated at rated voltage, because

$$n = (1/c_E)\,(V - Ir_a)/\Phi$$

$$n_0 = (1/c_E)\,(V/\Phi)$$

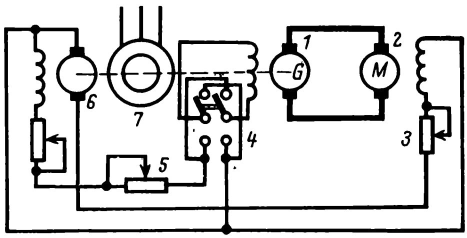


Fig. 4-28. Ward-Leonard system

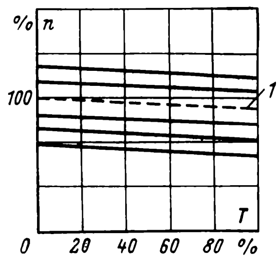


Fig. 4-29. Mechanical characteristics of the drive motor in the Ward-Leonard system

If we consecutively reduce the generator voltage, $V_g = V_m$, we shall obtain a family of $n = f(T)$ curves lying below curve *1*, but if we raise the exciting current of the motor, the resultant curves will lie above the curve.

4-17. Series- and Compound-Wound Motors

Series-wound motors find use as car starters, in some crane installations, and in electric traction. All of these applications permit a considerable variation in motor rpm as the static moment on the shaft is varied. The $T = f(P_2)$ and $n = f(P_2)$ characteristics, at $V =$ constant, for such a motor are shown in Fig. 4-30, and its connection in circuit, in Fig. 4-31.

The field winding of the motor is connected in series with the armature winding, and its resistance r_s is about the same as r_a, that is, very small. Therefore, a starting rheostat is as essential in this case as it is in other d.c. motors.

In operation, the motor current is

$$I = (V - E)/(r_a + r_s) \tag{4-21}$$

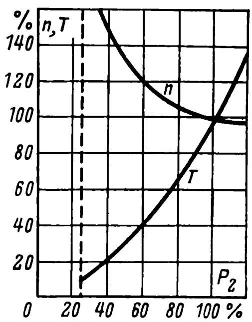


Fig. 4-30. Speed and torque characteristics of a series-wound motor

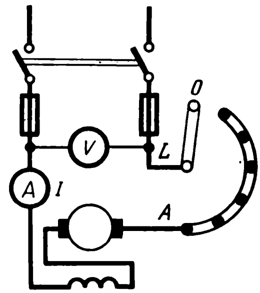


Fig. 4-31. Circuit of a series-wound motor

At low values of I (about 25 to 50% of the rated value), the flux in the motor is proportional to the current, $\Phi \sim I$, and

$$T = c_T I\Phi = c'_T II = c'_T I^2 \tag{4-22}$$

or, in words, *the torque is proportional to the current squared.* At high values of load on the shaft, I is very close to I_n, the motor core reaches saturation, and the torque becomes proportional to I.

The rpm

$$n = (1/c_E)\,\frac{V - I\,(r_a + r_s)}{\Phi} \tag{4-23}$$

drops drastically as the load increases, because Φ and $I\,(r_a + r_s)$ increase both at the same time. This form of response is called a *drooping* characteristic.

At a load less than 25 to 30% P_{2n}, the motor would run at an excessively high speed because the flux would be very small. This form of duty must be guarded against, because it could cause serious mechanical damage to the motor.

A compound-wound motor, that is, one with a series- and a shunt-field winding located on the main poles, behaves partly as a shunt-wound motor and partly as a series-wound motor. The two windings are connected so that their mmfs and, as a consequence, their respective fluxes, Φ_{sh} and Φ_s, are combined. This connection is known as series aiding.

In this case,

$$n = (1/c_E)\,\frac{V - I_a\,(r_a + r_s)}{\Phi_{sh} + \Phi_s} \tag{4-24}$$

$$T = c_T I_a\,(\Phi_{sh} + \Phi_s) \tag{4-25}$$

This type of motor is used in applications where the speed characteristic must be flatter than that of a series-wound motor. Compound-wound motors are employed as part of electric drives incorporating flywheels (shears, presses, and the like). The energy stored by the flywheel is expended only when the driven mechanism slows down. Under an impact load, a motor having a drooping characteristic tends to slow down, owing to which the kinetic energy stored by the flywheel is imparted to the driven mechanism. Provision of a flywheel makes it possible to use motors of lower power ratings.

Another use for the series field winding is in shunt-wound motors with a small air gap between the armature and poles and with a high armature current per unit length of armature circumference. In such motors, the demagnetizing effect of armature reaction affects the speed more than the voltage drop across the armature, Eq. (4-19). As load increases, the speed of such a motor rises rather than falls (Fig. 4-26, curve *2*), which is undesirable. If, however, the motor is fitted with another series (stabilizing) winding, its operation becomes more stable.

4-18. Losses and Efficiency

Some of the energy applied to an electric machine cannot be utilized to advantage, because it is dissipated as heat to the surroundings. This constitutes losses, of which there are several forms.

Iron loss, P_i. This form of loss occurs in the armature core and pole pieces as an outcome of cyclic (reversal of) magnetization and manifests itself as hysteresis loss and eddy-current loss. This power loss depends on the frequency at which reversals of magnetization occur, that is, $f = pn/60$, and the maximum value of magnetic induction, B_m.

Mechanical loss, P_m. This occurs in the bearings, due to friction between the brushes and commutator, and windage

(friction between air and the rotating parts). Mechanical power loss is proportional to the rpm, n. If the rpm and exciting current I_f are held constant, the sum of P_i and P_m is constant. It is called no-load power loss, $P_{n\text{-}l}$.

Electric loss, P_e. This is caused by the flow of current in the armature winding, contact resistance between the brushes and commutator, and the flow of current in the field and compole windings

$$P_e = I_a^2 r_a + P_b + I_a^2 r_{cp} + I_a^2 r_s + V_f I_f$$

Brush-contact loss, $P_b = \Delta V_b I_a$, is seen to be proportional to the voltage drop ΔV_b, which is assumed to be 2 V for carbon, graphite and graphitized brushes, and 0.6 V for copper-graphite brushes.

Stray power loss, P_s. This form of loss occurs in the armature winding and iron core owing to the field distortion caused by armature reaction and stray fields established around the coils being commutated. This loss component is estimated to be 0.01 to 0.005 $V_n I_n$ and is deemed proportional to I_a^2.

The efficiency of an electric machine is the ratio of useful power P_2 to the total power, P_1, applied. Then, for a generator

$$\eta_g = P_2/P_1 = \frac{VI}{VI + (P_i + P_m + P_e + P_s)} \times 100\% \qquad (4\text{-}26)$$

and for a motor

$$\eta_m = P_2/P_1 = \frac{VI - (P_i + P_m + P_e + P_s)}{VI} \times 100\% \qquad (4\text{-}27)$$

A plot of efficiency as a function of useful power P_2 is shown in Fig. 4-27. When the useful power is low and no-load loss $P_{n\text{-}l}$ is comparable with it, the efficiency is low. As the useful power rises, the efficiency rapidly improves, but no-load loss remains constant. As load increases, electric loss P_e increases in proportion to the current squared, and the rise of efficiency slows down. The efficiency is usually a maximum at 75 to 100% P_n and amounts to 70 to 93%. The higher values apply to larger machines.

Chapter Five   Alternating Current — Basic Concepts And Definitions

5-1. Alternating Current

Alternating current has today gained universal acceptance because it can be generated at any voltage — high for transmission to distant places and low for transmission over short distances, it can be used for various loads, requires simple single- and three-phase generators and motors which are reliable in operation, convenient to use and maintain, and show good performance. The few exceptions are several fields of technology, such as electrochemistry and electric traction, where use is predominantly made of direct current derived from a.c. by rectification.

The term 'alternating current' is usually applied to a periodic current which goes through a succession of values in a cyclic manner and in equal time intervals called the *period*, T. The number of such cycles per second is called the *frequency*, f. The frequency of an alternating current is the reciprocal of its period

$$f = 1/T \tag{5-1}$$

The unit of frequency is the *hertz*, abbreviated Hz

$$1\ \mathrm{Hz} = \mathrm{s}^{-1}$$

In the Soviet Union, the standard commercial (or power) frequency is 50 Hz. The frequencies used in electrothermal processes extend from 50 Hz to 50 MHz (1 MHz = $= 10^6$ Hz), and those in radio, from 10^5 to 10^{10} Hz.

The waveform of an alternating current is shown in Fig. 5-1, with time laid off as abscissa and current as ordinate.

The values that an alternating current, voltage or emf have at arbitrary instants of time t are called their instan-

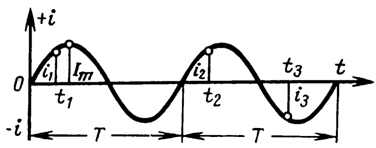


Fig. 5-1. Waveform of an alternating current

taneous values. They are designated by the low-case letters i, v and e, respectively.

The maximum value that a periodic quantity attains over a period is called its *amplitude* or *peak value*. It is symbolized by the respective capital letter with the subscript 'm', for example, I_m, V_m and E_m (see Fig. 5-1).

5-2. Generation of a Sinusoidal EMF

In engineering, use is mainly made of currents, voltages and emfs which obey the simple harmonic (sinusoidal) law. The choice of a sinusoidal waveform for electrical quantities is dictated by technical and economic considerations — with this waveform electric machines and apparatus have better efficiency owing to lower losses and show better performance, the insulation operates under easier conditions, and calculations are far simpler to make.

The elementary generator shown in Fig. 5-2 has a magnet with poles labelled N and S, between which is rotating a cylindrical armature, A, built up of electrical-sheet-steel laminations or punchings. On its surface, the armature carries a coil of w turns, whose leads are connected via slip rings and brushes to an external circuit.

The poles are shaped in such a way that the magnetic induction B in the air gap all the way round the armature varies sinusoidally

$$B = B_m \sin \alpha$$

where α is the angle that the neutral plane $00'$ makes with the plane of the coil.

When the armature is rotating at an angular velocity $\Omega = d\alpha/dt$, an emf is induced in each active side of the coil (Fig. 5-3), given by

$$e' = Blv = B_m lv \sin \alpha = B_m lv \sin \Omega t \tag{5-2}$$

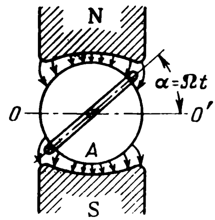


Fig. 5-2. A.C. generator in sketch form

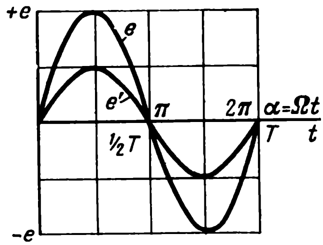


Fig. 5-3. Waveforms of alternating emfs

Because the coil has w turns, and the number of coil sides is $2w$, the emf induced in the armature winding is

$$e = e' \times 2w = 2B_m wlv \sin \Omega t = E_m \sin \Omega t \quad (5\text{-}3)$$

where $E_m = 2B_m wlv$ is the amplitude of the emf.

Initially, that is, at $t = 0$, the plane of the coil coincides with the neutral plane, because $\alpha = \Omega t = 0$.

In a generator having one pair of poles ($p = 1$), one revolution of the armature, $\alpha = 2\pi$, corresponds to one alternation of emf. When the armature is driven at constant speed, the angular velocity Ω is likewise constant

$$\Omega = \alpha/t = 2\pi/T = 2\pi f = \omega \quad (5\text{-}4)$$

In a generator having p pairs of poles (Fig. 5-4), each active conductor of the armature winding passes under p pairs of poles during every revolution, so one revolution corresponds to p alternations. Hence,

$$e = E_m \sin (p\alpha) = E_m \sin (p\Omega t) = E_m \sin \omega t \quad (5\text{-}5)$$

A plot of e as a function of t, $e = f(t)$, during one revolution of the armature in a generator having $p = 2$ is shown in Fig. 5-5.

The product $p\alpha$ is called the electric angle. The ratio of the electric angle to the time during which it changes is called the *electric angular velocity* or *angular frequency*. Obviously,

$$\omega = p\Omega = p\alpha/t = p \times 2\pi/pT = 2\pi f \quad (5\text{-}6)$$

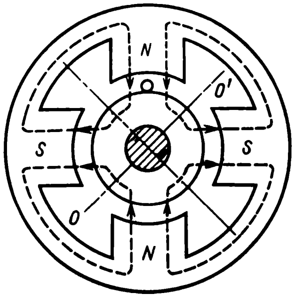


Fig. 5-4. A. C. generator with two pairs of poles

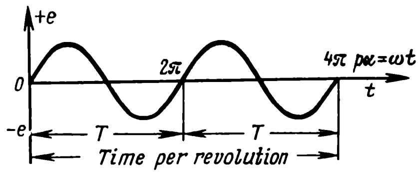


Fig. 5-5. Waveform of the alternating emf produced by a generator with two pairs of poles

If the armature is running at n revolutions per minute, its speed is $n/60$ rpm, and its frequency is

$$f = np/60 \qquad (5\text{-}7)$$

Example 5-1. Given: A generator with two pairs of poles ($p=2$), rotating at 1500 rpm. To find: The frequency of the alternating current produced by the generator.

Solution.

$$f = pn/60 = 2 \times 1500 \div 60 = 50 \text{ Hz}$$

Example 5-2. Given: A hydroalternator with a rated speed of 250 rpm and a frequency of 50 Hz. To find: The number of poles that the generator has.

Solution.

$$p = f \times 60/n = 50 \times 60 \div 250 = 12 \text{ pairs}$$

5-3. Phase Difference

Imagine a generator in which the armature carries two identical loops, *1* and *2*, spaced some distance apart (Fig. 5-6). When the armature is rotated, emfs are induced in the loops, having the same frequency and the same amplitude, because the loops are rotating at the same angular velocity in the same magnetic field. However, the loops are shifted in space from each other, so they pass under the

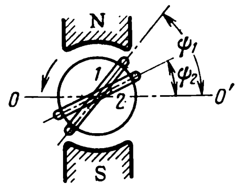


Fig. 5-6. Generator armature carrying two loops

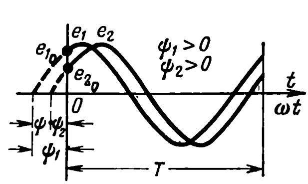


Fig. 5-7. Waveforms of two alternating emfs

centre of each pole at different times, and their emfs reach their amplitude (peak) values likewise at different times.

Suppose that the armature is rotating at an angular velocity ω counter-clockwise. Then, initially (at time $t = 0$), the loops will make angles ψ_1 and ψ_2 with the neutral plane 00′ (see Fig. 5-6). Accordingly, the emfs induced in the loops will be

$$e_1 = E_m \sin(\omega t + \psi_1)$$

and (5-8)

$$e_2 = E_m \sin(\omega t + \psi_2)$$

where the angle $(\omega t + \psi)$ is called the phase angle or simply the phase of a sinusoidal quantity. As is seen, the instantaneous value of a sinusoidal quantity is proportional to its amplitude and phase.

Let us have a closer look at the plots of the sinusoidal emfs in Fig. 5-7. Initially, that is, at $t = 0$, the emfs induced in the loops are

$$e_{1_0} = E_m \sin \psi_1$$

and

$$e_{2_0} = E_m \sin \psi_2$$

In Fig. 5-7 they are represented by the initial (zero-time) ordinates. The phase angles ψ_1 and ψ_2 determining the emfs at the initial instants of time are called the *initial phases*, or *epochs*, of the emfs.

To sum up, *a sinusoidal quantity can be completely defined by specifying* (1) *its amplitude*, (2) *its frequency or period, and* (3) *its initial phase or epoch.*

The difference in epoch between two sinusoidal quantities of the same frequency is called their *phase difference*

$$\psi = \psi_1 - \psi_2 \tag{5-9}$$

The phase difference indicates how much, in terms of period fraction or time, $t = \psi/\omega = \psi T/2\pi$, one sinusoidal quantity leads the other in reaching the start of a cycle or period.

The start of a cycle or period is the time at which the sinusoidal quantity of interest passes through a zero value (makes a zero crossing, as it is sometimes stated) from negative to positive. The quantity reaching this instant ahead of the other is said to be leading in phase, and the other is said to be lagging in phase.

Two sinusoidal quantities having the same epochs are said *to be in phase*. Two sinusoidal quantities having a phase difference of 180° are said *to be in anti-phase*.

Example 5-3. Given: Two emfs defined by

$$e_1 = E_m \sin(\omega t + 60°) \quad \text{and} \quad e_2 = E_m \sin(\omega t + 30°)$$

To find: The phase difference between e_1 and e_2 and the time difference, if the frequency is 50 Hz.

Solution.

Let us write the epochs ψ_1 and ψ_2 in radians

$$\psi_1 = 60 \times 2\pi/360 = \pi/3$$

$$\psi_2 = 30 \times 2\pi/360 = \pi/6$$

Hence, the phase difference is

$$\psi = \psi_1 - \psi_2 = \pi/3 - \pi/6 = \pi/6$$

The period is

$$T = 1/f = 1/50 = 0.02 \text{ s}$$

so the time difference between e_1 and e_2 is

$$t = \psi/\omega = \pi T/(6 \times 2\pi) = T/12 = 0.00166 \text{ s}$$

5-4. Root-Mean-Square Values of Current and Voltage

When solving a.c. circuits, we ordinarily use the root-mean-square (rms) values of alternating current, voltage and emf.

The rms values of current, voltage and emf are symbolized by the capital letters I, V and E, respectively.

Usually, it is the rms values that are stated on the dials of measuring instruments and in data sheets.

The rms value is that value of an alternating current or voltage which produces the same heating effect as would be produced by an equal value of direct current or voltage on passing through the same resistance over a period.

The amount of heat that an alternating current produces in a resistance r over an infinitesimal time interval dt is

$$dw_h = i^2 r\, dt$$

and over a period (cycle),

$$W_h = \int_0^T dw_h = \int_0^T i^2 r\, dt$$

On equating the above expression for W_h to the amount of heat, $I^2 rT$, produced by a direct current I in the same resistance r over the same time interval T, we get

$$I^2 rT = \int_0^T i^2 r\, dt$$

On cancelling out the common term r, we obtain the rms value of current

$$I = \sqrt{\frac{1}{T}\int_0^T i^2\, dt} \tag{5-10}$$

Figure 5-8 shows the waveform of the instantaneous current i and the waveform of the instantaneous current squared, i^2. The area bounded by the latter waveform and the x-axis represents the quantity defined by the integral $\int_0^T i^2 dt$, drawn to a definite scale. The height Ab of the rectangle $AbcEA$ whose area is equal to that bounded by the i^2 curve ($ABCDE$) and the x-axis is also equal to the mean ordinate of the i^2 curve and represents the rms current squared, that is I^2.

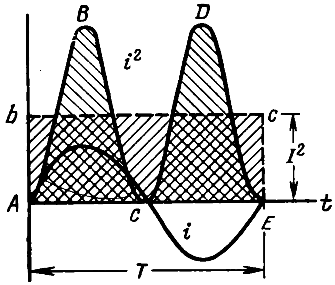


Fig. 5-8. Waveforms of an alternating current and a squared current

If variations in a current obey the sinusoidal law

$$i = I_m \sin \omega t$$

then

$$I = \sqrt{\frac{1}{T} \int_0^T I_m^2 \sin^2 \omega t \, dt}$$

Since

$$\int_0^T \sin^2 \omega t \, dt = \int_0^T \frac{1 - \cos 2\omega t}{2} \, dt$$

$$= \frac{1}{2} \int_0^T dt - \frac{1}{2} \int_0^T \cos 2\omega t \, dt$$

$$= T/2 - 0 = T/2$$

then

$$I = \sqrt{\frac{1}{T} I_m^2 \frac{T}{2}} = \sqrt{I_m^2/2} = I_m/\sqrt{2} = \text{(approx.) } 0.707 \ I_m \tag{5-11}$$

Similarly, for the rms values of sinusoidal voltages and emfs we may write

$$V = V_m/\sqrt{2} = \text{(approx.) } 0.707 \ V_m \tag{5-12}$$

and

$$E = E_m/\sqrt{2} = \text{(approx.) } 0.707 \ E_m$$

Example 5-4. Given: The voltage measured by a voltmeter is $V = 220$ V. To find: The voltage amplitude.
Solution.

$$V_m = V\sqrt{2} = 220 \times 1.41 = 310\,V$$

In addition to the rms values of current and voltage, use is sometimes made of the *average value* of current and voltage.

The average value of a sinusoidal current over a period is zero, because the same quantity of electricity Q passes through the cross-sectional area of the conductor in either direction during the first and second half-period. As a consequence, the quantity of electricity passing through the cross-sectional area of the conductor over a period is zero, and so is the value of sinusoidal current averaged over a period. This is why the average value of a sinusoidal current, I_{av}, is that taken over a half-period (half-cycle), during which the current remains positive.

The average current is the ratio of the quantity of electricity passing through the cross-sectional area of a conductor over a half-period (half-cycle) to the duration of the half-cycle, namely

$$I_{av} = Q/(T/2) = 2Q/T = \frac{2}{T}\int_0^{T/2} i\,dt \tag{5-13}$$

In taking the integral, the reference time, $t = 0$, must be chosen so that it coincides with the start of a cycle. The average values of voltage and emf are

$$V_{av} = \frac{2}{T}\int_0^{T/2} v\,dt$$

and

$$E_{av} = \frac{2}{T}\int_0^{T/2} e\,dt \tag{5-14}$$

The value of current averaged over a half-cycle can graphically be depicted as the height of a rectangle whose base is equal to $T/2$ and whose area is equal to that bounded by the x-axis and the current waveform between the start and the half of a cycle (Fig. 5-9).

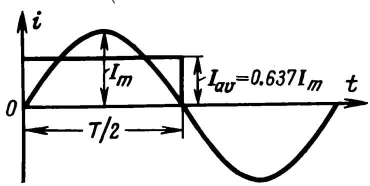


Fig. 5-9. Current averaged over a half-cycle

The average value of a sinusoidal current can be expressed in terms of its amplitude (peak) value as

$$I_{av}=\frac{2}{T}\int_0^{T/2} i\,dt=\frac{2I_m}{T}\int_0^{T/2}\sin\omega t\,dt$$

$$=-\frac{2I_m}{T\omega}\cos\omega t\Big|_0^{T/2}=2I_m/\pi=0.637\,I_m \quad (5\text{-}15)$$

The same relation holds for voltages and emfs:

$$V_{av}=2V_m/\pi$$

and (5-16)

$$E_{av}=2E_m/\pi$$

5-5. Vector Diagrams

Sinusoidal quantities are represented graphically by a class of curves known as sinusoids (see Secs. 5-1 through 5-3) or by rotating vectors. The latter method appreciably simplifies the graphical representation of sinusoidal quantities and the graphical determination of their sums and differences.

When a sinusoidal quantity, say, an emf defined by

$$e=E_m\sin(\omega t+\psi)$$

is represented by a rotating vector (Fig. 5-10), the vector is drawn so that its lenght OA represents the amplitude value E_m on a certain scale. The angle between the vector and the positive x-axis at time $t=0$ gives the initial phase or epoch ψ, and the angular velocity of the vector is equal to the angular frequency ω of the quantity. A pro-

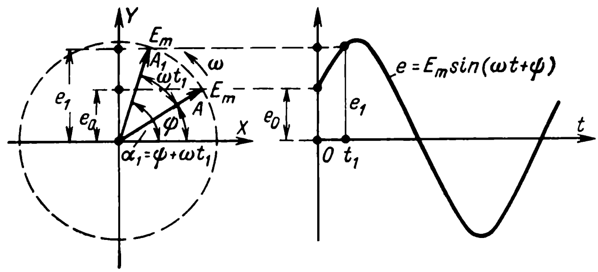


Fig. 5-10. Representation of a sinusoidal quantity by a rotating vector

jection of the vector on the y-axis to the same scale then gives the instantaneous value of the emf.

To demonstrate, let at time $t = 0$ the emf $e_0 = E_m \sin \psi$ be represented by a projection of the vector OA on the y-axis. At time t_1, the emf $e_1 = E_m \sin (\omega t_1 + \psi)$ is represented by a projection of the same vector, but in a new position, OA_1, onto the y-axis. A set of several vectors representing sinusoidal quantities of the same frequency is referred to as a *vector diagram*.

Because the angular velocity of all vectors in the same diagram is also the same, their relative position does not change with time. The reference time for a periodic waveform can be chosen at will, so in plotting a vector diagram one of the vectors may be positioned arbitrarily, and the remaining vectors will then be drawn in their proper relative positions with regard to the datum or reference vector so as to maintain the actual phase differences unchanged.

The addition of two sinusoidal quantities can be replaced by the addition of vectors, each of which represents a particular sinusoidal quantity. Let, for example, there be two emfs

$$e_1 = E_{m1} \sin (\omega t + \psi_1)$$

and

$$e_2 = E_{m2} \sin (\omega t + \psi_2)$$

which are represented by the vectors OA and OB in the diagram of Fig. 5-11. In order to combine them, the vector

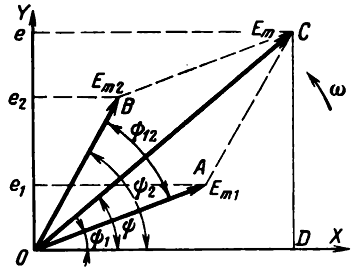


Fig. 5-11. Vectorial addition of two emfs

OA is shifted parallel to itself so that its starting point coincides with the terminal point (or terminus) of the vector OB. Then the closing vector OC will represent the total emf, e. The validity of this statement is borne out by the fact that the projections of the vectors OA and OB into the y-axis represent the instantaneous values of e_1 and e_2, and the sum of the projections is equal to the vector OC representing the total emf.

The vector triangle can be used to find the amplitude of the total emf and the tangent of its epoch angle (see Fig. 5-11).

The subtraction of sinusoidal quantities is performed as the addition of the minuend and the subtrahend taken with the opposite sign, namely

$$e_1 - e_2 = e_1 + (-e_2)$$

or

$$\bar{E}_{m1} - \bar{E}_{m2} = \bar{E}_{m1} + (-\bar{E}_{m2})$$

Chapter Six — Single-Phase A. C. Circuits

6-1. A General Outline of A.C. Circuits

When a direct voltage is impressed on an electric circuit, the power and energy stored up by its electric and magnetic fields remain constant.

When an alternating voltage is applied to a circuit, the two quantities vary with time.

Any electric circuit or network which converts electricity to heat and in which the energy stored by its electric and magnetic fields is subject to variations can be completely specified by three parameters, namely, resistance r, inductance L, and capacitance C.

In engineering, one encounters circuits controlled by one of the three parameters, that is, r, L or C, to the exclusion of the other two. For example, an incandescent lamp, a resistor or a heating appliance can be described in terms of only resistance r; transformers at no-load in terms of only inductance L; cables at no-load in terms of only capacitance C.

6-2. A Circuit Containing Only a Resistance

(a) Voltage and Current

When a sinusoidal voltage

$$v = V_m \sin \omega t$$

is applied to a circuit of resistance r (Fig. 6-1), the current in the circuit is, by Ohm's law, given by

$$i = v/r = (V_m/r) \sin \omega t = I_m \sin \omega t$$

The above expression states that the *current in a resistance r varies sinusoidally, being in phase with the applied voltage* (Fig. 6-2*a* and Fig. 6-3).

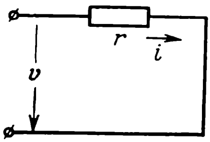
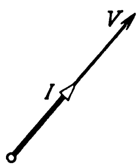


Fig. 6-1. Circuit containing only a resistance



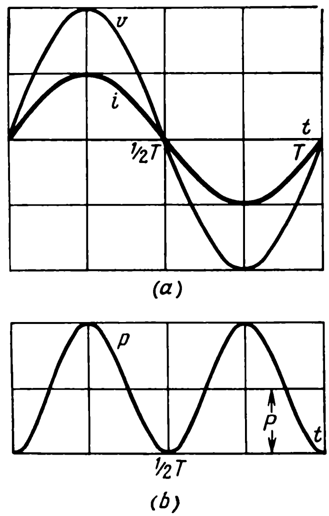


Fig. 6-2. Waveforms of current, voltage and power for a circuit containing only a resistance

Fig. 6-3. Vector diagram of a circuit containing only a resistance

For a specified circuit, Ohm's law remains valid for both instantaneous and amplitude values, $i = v/r$ and $I_m = V_m/r$, and rms values

$$I = 0.707\ I_m = 0.707\ V_m/r = V/r \tag{6-1}$$

(b) Power

The instantaneous value of power or simply *instantaneous power* is the product of the voltage and current taken at the same instance, that is,

$$p = vi = i^2r = I^2{}_m r \sin^2 \omega t$$

Electric energy is converted to heat irrespective of the direction in which current flows, so the instantaneous power is positive for the forward and reverse directions of current flow (see Fig. 6-2b).

Since

$$\sin^2 \omega t = 1/2 - 1/2 \cos 2\omega t$$

and

$$1/2 I^2_m = (I_m/\sqrt{2})^2 = I^2$$

we may write

$$p = I^2_m r \sin^2 \omega t = {}^1/_2\, I^2_m r - {}^1/_2\, I^2_m r \cos 2\omega t$$
$$= I^2 r - I^2 r \cos 2\omega t$$

In the above expression, $I^2 r$ is the direct component which represents the average rate of converting electricity to heat or, which is the same, the power averaged over a period

$$P = I^2 r = IIr = IV \qquad (6\text{-}2)$$

It is called the *active power*. An alternative name is the *resistive power*, because it is dissipated in the resistance.

As in d.c. circuits, active power is measured in watts, W.

6-3. A Circuit Containing Only an Inductance

(a) Voltage and Current

When a circuit containing only an inductance (Fig. 6-4) is traversed by a current $i = I_m \sin \omega t$, an emf of self-induction defined by Eq. (3-37) is induced in it

$$e_L = -L\, di/dt = -L\, d\,(I_m \sin \omega t)/dt$$
$$= -LI_m \omega \cos \omega t$$
$$= E_{Lm} \sin (\omega t - \pi/2) \qquad (6\text{-}3)$$

If an inductance-containing circuit has a negligible resistance ($r = 0$), then by Kirchhoff's voltage (second) law

$$v + e_L = ir = 0$$

Hence, the voltage across the terminals of the circuit is

$$v = -e_L = LI_m \omega \cos \omega t = V_m \sin (\omega t + \pi/2) \quad (6\text{-}4)$$

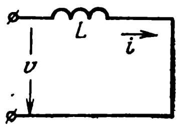


Fig. 6-4. Circuit containing only an inductance

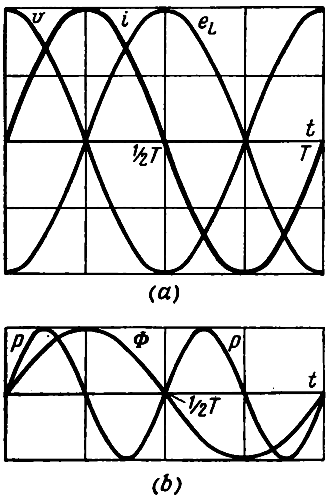


Fig. 6-5. Waveforms of current, voltage and power for a circuit containing only an inductance

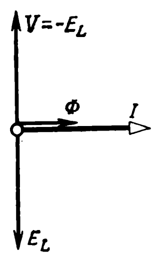


Fig. 6-6. Vector diagram for a circuit containing only an inductance

Or, in words, the applied voltage produces in the circuit a current whose magnetic field, as it varies, induces at every instant an emf of self-induction which is equal in magnitude but opposite in direction to the applied voltage; that is, the induced emf balances out the applied voltage (Figs. 6-5*a* and *b* and 6-6).

It follows from Eq. (6-4) and Figs. (6-5) and (6-6) that the current is lagging behind the voltage by a quarter of a period ($\pi/2$) or is leading the emf e_L by a quarter of a period in phase. This is because the induced emf, e_L, is proportional to the time rate of change of current. As the current passes through its maximum value, its rate of change is zero, and the induced emf is likewise equal to zero.

When the current makes a zero-crossing (passes through its zero value), its rate of change is a maximum, and the induced emf is $e_L = E_{Lm}$. By Lenz's law, the emf e_L opposes the current in the case of a positive increment ($di/dt > 0$); conversely, e_L aids the current in the case of a negative increment ($di/dt < 0$). This explains why, for example, during the first quarter of a cycle (see Fig. 6-5), when the current is rising, the emf e_L is negative, whereas it is positive during the second quarter of a cycle, when the current is decreasing.

(b) Inductive Reactance

From Eqs. (6-3) and (6-4) it follows that

$$V_m = E_{Lm} = I_m \omega L \tag{6-5}$$

Hence, Ohm's law for amplitudes may be written as

$$I_m = V_m/\omega L = V_m/x_L, \tag{6-6}$$

Dividing the above expression by $\sqrt{2}$ gives Ohm's law for rms values

$$I = V/\omega L = V/x_L \tag{6-7}$$

The ratio of the applied voltage to the current in the circuit

$$V/I = x_L = \omega L = 2\pi f L \tag{6-8}$$

is called the *inductive reactance of the circuit*. It is proportional to the circuit inductance and the current frequency. In d.c. circuits, it reduces to zero.

(c) Power

The instantaneous power in an inductive circuit is

$$\begin{aligned} p &= vi = V_m \sin(\omega t + \pi/2)\, I_m \sin \omega t \\ &= V_m I_m \cos \omega t \sin \omega t \\ &= V_m I_m \, {}^1/_2 \sin 2\omega t \\ &= VI \sin 2\omega t \end{aligned} \tag{6-9}$$

As is seen, the power varies at twice the supply frequency (see Fig. 6-5), reaching a positive peak, $VI = I^2 \omega L$, and a negative peak twice every cycle, or period.

Irrespective of the direction in which the current flows, a rise in the current and, as a consequence, in the magnetic flux (the first and third quarters of a cycle, Fig. 6-5), causes the energy stored by the magnetic field to build up from zero to a maximum, Eq. (3-39)

$$W_m = LI_m^2/2 = LI^2$$

This energy is supplied by the associated generator, and the circuit is operating as a load, which corresponds to the positive value of the power in the circuit.

A decrease in the current and, as a consequence, in the magnetic flux (the second and fourth quarters of a cycle) brings about a decrease in the energy stored by the magnetic field from a maximum to zero, and the energy is returned to the generator. Now the circuit is operating as a source of power, which corresponds to the negative value of the power in the circuit.

The active power P in an inductive circuit is zero. The maximum value of power in an inductive circuit has come to be known as the *reactive power*, symbolized by Q.

From Eq. (6-9) it follows that

$$Q = V_m I_m/2 = VI = I^2 \omega L = \omega W_m \qquad (6\text{-}10)$$

Reactive power is measured in reactive volt-amperes, the unit being the *volt-ampere reactive*, VAr.

Example 6-1. Given: A circuit containing an inductance of 0.02 H energized with 127 V at 50 Hz.

To find: 1. The inductive reactance, current, and reactive power of the circuit.

2. The inductive reactance and current of the circuit at 1000 Hz.

Solution.

1. $x_L = 2\pi f L = 2\pi \times 50 \times 0.02 = 6.28$ ohms

$I = V/x_L = 127/6.28 = 20.25$ A

$Q = VI = 127 \times 20.25 =$ (approx.) 2572 VAr

2. $x_L = 2\pi f L = 2\pi \times 1000 \times 0.02 = 125.6$ ohms

$I = V/x_L = 127/125.6 =$ (approx.) 1.01 A

(d) The Voltage Across an Inductance as a Function of the Magnetic Flux

In some cases, when solving an a.c. circuit, it is convenient to express the voltage across an inductance in terms of magnetic flux.

If all turns of a coil (loop) link the same magnetic flux, the amplitude of the flux linkage of self-induction is

$$\Psi_m = w\Phi_m = LI_m$$

In this case, the emf of self-induction and the terminal voltage are equal

$$V = E_L = \omega LI_m/\sqrt{2} = 2\pi fw\Phi_m/\sqrt{2} = 4.44fw\Phi_m \quad (6\text{-}11)$$

6-4. A Circuit Containing a Resistance and an Inductance

(a) Voltage and Current

If an inductor of resistance r and of inductance L (Fig. 6-7) is traversed by an alternating current (Figs. 6-8 and 6-9)

$$i = I_m \sin \omega t$$

then by Kirchhoff's voltage (second) law

$$v + e_L = ir$$

Hence, the terminal votage across the coil (circuit) is

$$v = ir - e_L = ir + L\,di/dt = v_a + v_L$$

Here, the first term, $v_a = ir$, is called the *active or resistive voltage* and the second term, $v_L = -e_L = L\,di/dt$ is called the *reactive* or *inductive voltage.*

The active voltage (see Figs. 6-8 and 6-9)

$$v_a = ir = I_m r \sin \omega t = V_{a,m} \sin \omega t$$

varies sinusoidally and is in phase with the current.

The amplitude of the active voltage is

$$V_{a,m} = I_m r$$

and its rms value is

$$V_a = Ir$$

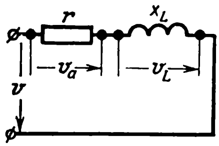


Fig. 6-7. Circuit containing a resistance and an inductance

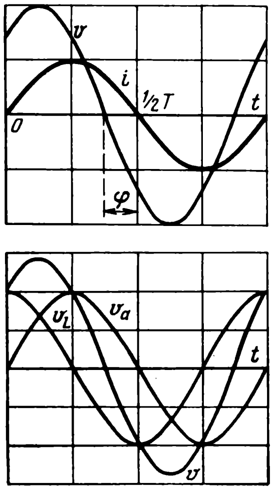


Fig. 6-8. Waveforms of current and voltage for a circuit containing a resistance and an inductance

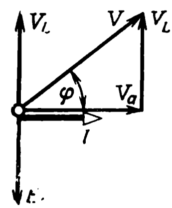


Fig. 6-9. Vector diagram for a circuit containing a resistance and an inductance

The reactive voltage (see Figs. 6-8 and 6-9)

$$v_L = L\,di/dt = \omega L I_m \cos \omega t = V_{Lm} \sin(\omega t + \pi/2)$$

varies sinusoidally and leads the current by 90° (it is said to be in quadrature leading with the current).

The amplitude of reactive voltage is

$$V_{L,m} = \omega L I_m$$

and its rms value is

$$V_L = \omega\, L I = x_L I$$

The voltage across the terminals of the circuit shown in Figs. 6-8 and 6-9

$$v = v_a + v_L = V_{a,m} \sin \omega t + V_{Lm} \sin (\omega t + \pi/2)$$
$$= V_m \sin (\omega t + \varphi) \qquad (6\text{-}12)$$

varies sinusoidally and leads the current by an angle φ in phase.

The vectors V_a, V_L and V (see Fig. 6-9) form a right-angled *voltage triangle* from which it follows that

$$V = \sqrt{V_a^2 + V_L^2} \qquad (6\text{-}13)$$

This relation also applies to the amplitudes of the respective voltages

$$V_m = \sqrt{V_{a,m}^2 + V_{Lm}^2}$$

The phase difference between the voltage across and the current in an a.c. circuit can be found from the voltage triangle of Fig. 6-9 as

$$\cos \varphi = V_a/V \text{ or } \tan \varphi = V_L/V_a$$

(b) The Impedance of a Circuit

On expressing the component voltages in Eq. (6-13) as the products of currents, reactances and resistances, we obtain

$$V = \sqrt{(Ir)^2 + (Ix_L)^2} = I \sqrt{r^2 + x_L^2} = Iz$$

Hence the current in the circuit is

$$I = V/z = V/\sqrt{r^2 + x_L^2} \qquad (6\text{-}14)$$

The above equations state Ohm's law for a circuit of resistance r and inductance L in terms of the rms values of current and voltage.

The quantity

$$z = \sqrt{r^2 + x_L^2} = \sqrt{r^2 + (\omega L)^2} \qquad (6\text{-}15)$$

is termed the *impedance* of a circuit.

Graphically, the resistance r, the inductive reactance x_L and the impedance z are depicted as the sides of a right-angled *impedance triangle* (Fig. 6-10). This triangle can be derived from the voltage triangle by reducing its sides to $1/I$ of their former length.

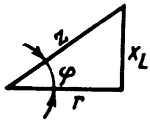


Fig. 6-10. Impedance triangle for a circuit containing a resistance and an inductance

The angle between the sides z and r gives the phase difference φ between the voltage and current, because

$$\cos\varphi = V_a/V = Ir/Iz = r/z$$

and

$$\tan\varphi = V_L/V_a = x_L/r \tag{6-16}$$

The greater the reactive voltage in comparison with the active voltage or the greater the inductive reactance in comparison with the resistance, the greater the angle by which the current lags in phase behind the voltage.

(c) Power

In a circuit containing a resistance r and an inductance L the instantaneous power is

$$\begin{aligned} p &= vi = V_m \sin(\omega t + \varphi)\, I_m \sin \omega t \\ &= (V_m I_m/2) \cos\varphi - (V_m I_m/2) \cos(2\omega t + \varphi) \\ &= VI \cos\varphi - VI \cos(2\omega t + \alpha) \end{aligned} \tag{6-17}$$

It follows from Eq. (6-17) that the instantaneous power in a circuit is the sum of a direct component, $VI \cos\varphi$, and an alternating component, $VI \cos(2\omega t + \varphi)$, varying sinusoidally at twice the supply frequency. The power averaged over a period, ordinarily used in calculations, is equal to the direct power $VI \cos\varphi$, because a harmonic function averaged over a cycle is equal to zero.

Thus, the average power in a circuit is equal to the product of the rms voltage and rms current multiplied by $\cos\varphi$, or, mathematically,

$$P = VI \cos\varphi \tag{6-18}$$

Recalling that

$$V \cos\varphi = V_a = Ir$$

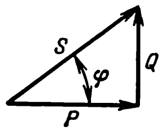


Fig. 6-11. Power triangle

we get

$$P = VI \cos \varphi = V_a I = I^2 r$$

As is seen, the average power in a resistance, given by Eq. (6-2) to be $P = V_a I$, is at the same time the average or active power in a circuit containing both r and L, that is, $P = VI \cos \varphi$.

The reactive power of a circuit, defined by Eq. (6-10), characterizes the exchange of energy that takes place between a generator and the associated circuit

$$Q = V_L I = I^2 x_L = I^2 z \sin \varphi = VI \sin \varphi \qquad (6\text{-}19)$$

and is equal, as stated mathematically above, to the product of the rms voltage and current multiplied by sin φ.

The product of the rms voltage and current

$$S = VI \qquad (6\text{-}20)$$

is called the *apparent power* in a circuit. It is measured in *volt-amperes*, VA.

The active, reactive and apparent powers are graphically depicted as the sides of a right-angled *power triangle* (Fig. 6-11), because they are connected by a relation of the form

$$P^2 + Q^2 = S^2$$

or

$$(VI \cos \varphi)^2 + (VI \sin \varphi)^2 = (VI)^2 \qquad (6\text{-}21)$$

The power triangle can be derived from the voltage triangle if we multiply its sides by the circuit current.

The ratio of the active to the apparent power

$$P/S = \cos \varphi \qquad (6\text{-}22)$$

is called the *power factor*.

The overall dimensions, mass, cost and design of an electric machine or apparatus are all controlled by its rated apparent power, $S_n = V_n I_n$, and the apparent power

S characterizes the degree to which the machine or apparatus is utilized in a particular duty.

Example 6-2. Given: An inductor of $L = 102$ mH $= 0.102$ H and $r = 24$ ohms, energized with 240 V at 50 Hz.
To find: x_L, z, I, V_a, V_L, $\cos \varphi$, and P.
Solution.

$$x_L = 2\pi f L = 2\pi \times 50 \times 0.102 = 32 \text{ ohms}$$

$$z = \sqrt{r^2 + x_L^2} = \sqrt{24^2 + 32^2} = 40 \text{ ohms}$$

$$I = V/z = 240/40 = 6 \text{ A}$$

$$V_a = Ir = 6 \times 24 = 144 \text{ V}$$

$$V_L = Ix_L = 6 \times 32 = 192 \text{ V}$$

$$\cos \varphi = r/z = 24/40 = 0.6$$

$$P = VI \cos \varphi = 240 \times 6 \times 0.6 = 864 \text{ W}$$

6-5. A Series Circuit Containing Resistances and Inductances

The voltages across the resistances of two inductors connected in series (Fig. 6-12), $V_{a1} = Ir_1$ and $V_{a2} = Ir_2$, are in phase with the current I. The voltages across the inductive reactances of the inductors, $V_{L1} = Ix_{L1}$ and $V_{L2} = Ix_{L2}$, lead the current by 90° (they are said to be in quadrature leading) (Fig. 6-13).

The voltage across the terminals of a series circuit containing two inductors can be found from the voltage triangle

$$V = \sqrt{(V_{a1} + V_{a2})^2 + (V_{L1} + V_{L2})^2} = \sqrt{V_a^2 + V_L^2}$$

On expressing the component voltages in terms of currents, resistances and inductive reactances, we get

$$V = I\sqrt{(r_1 + r_2)^2 + (x_{L1} + x_{L2})^2} = I\sqrt{r^2 + x_L^2} = Iz$$

where

$r = r_1 + r_2$ is the resistance of the circuit

$x_L = x_{L1} + x_{L2}$ is the inductive reactance of the circuit.

The impedance of the circuit is

$$z = \sqrt{r^2 + x_L^2}$$

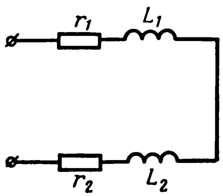


Fig. 6-12. Series circuit containing resistances and inductances (two inductors in series)

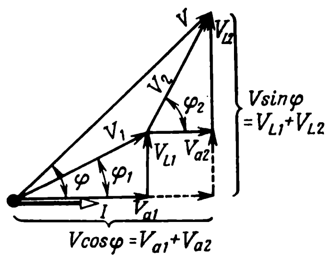


Fig. 6-13. Vector diagram for a series circuit

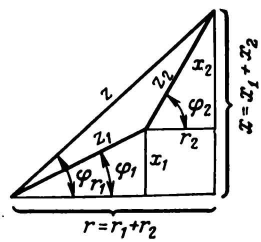


Fig. 6-14. Impedance triangle for a series circuit

In Fig. 6-14, the impedance is represented by the hypotenuze of a right-angled impedance triangle which can be derived from the voltage triangle, if we reduce each of its sides to $1/I$ of its original length.

The circuit current

$$I = V/z$$

lags behind the voltage across the circuit by an angle φ which can be determined from its cosine or tangent

$$\cos\varphi = r/z \text{ and } \tan\varphi = x_L/r$$

The average, or active, power in the circuit containing two inductors is

$$P = P_1 + P_2 = VI \cos\varphi$$

The reactive and apparent powers in the same circuit are

$$Q = VI \sin\varphi \text{ and } S = VI$$

6-6. A Parallel Circuit Containing Resistances and Inductances

The current in the first parallel path or branch (Fig. 6-15)

$$I_1 = V/z_1 = V/\sqrt{r_1^2 + x_{L1}^2}$$

lags behind the voltage in phase by an angle whose tangent is

$$\tan \varphi_1 = x_{L1}/r_1$$

The current in the second parallel path or branch (Fig. 6-15)

$$I_2 = V/z_2 = V/\sqrt{r_2^2 + x_{L2}^2}$$

lags behind the voltage in phase by an angle whose tangent is

$$\tan \varphi_2 = x_{L2}/r_2$$

To simplify the solution of parallel circuits, the current in each parallel path is resolved into two components. One component is the active current, I_a, that is, the current which is in phase with the applied voltage. The other component is the reactive current, I_r, that is, the current which is in quadrature with the voltage (displaced through 90° from it).

The component currents in the first parallel path (Fig. 6-16) are

$$\left.\begin{aligned} &I_{a1} = I_1 \cos \varphi_1 = (V/z_1)(r_1/z_1) = Vr_1/z_1^2 = Vg_1 \\ &\text{and} \\ &I_{r1} = I_1 \sin \varphi_1 = (V/z_1)(x_{L1}/z_1) = Vx_{L1}/z_1^2 = Vb_1 \end{aligned}\right\} \quad (6\text{-}23)$$

where g_1 is the conductance and b_1 is the susceptance of the circuit.

In constructing a vector diagram, the active current vector is laid off in the direction of the voltage vector. The inductive current component is laid off at right angles in the clockwise direction. The closing side of the current triangle is the current vector of the first path

$$I_1 = \sqrt{I_{a1}^2 + I_{r1}^2} = \sqrt{(Vg_1)^2 + (Vb_1)^2} = V\sqrt{g_1^2 + b_1^2} = Vy_1 \quad (6\text{-}23a)$$

where

$$y_1 = 1/z_1 = \sqrt{g_1^2 + b_1^2}$$

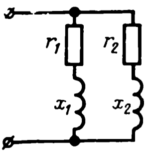


Fig. 6-15. Parallel circuit containing resistances and inductances (two inductors in parallel)

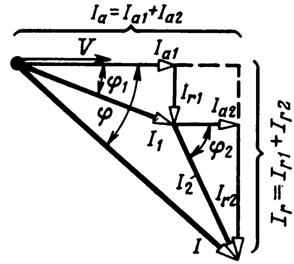


Fig. 6-16. Vector diagram for an a.c. parallel circuit

is the *admittance* of the first parallel path.

For the second parallel path

$$I_{a2}=I_2\cos\varphi_2$$

$$I_{r2}=I_2\sin\varphi_2$$

$$I=\sqrt{I_{a2}^2+I_{r2}^2}$$

The sum of the active currents in the two parallel paths having the same phase gives the active component of the circuit (total) current

$$I_a=I_{a1}+I_{a2}$$

The sum of the reactive currents in the two parallel paths, having the same phase, gives the reactive component of the circuit (total) current

$$I_r=I_{r1}+I_{r2}$$

The total current in the series part of the circuit is

$$I=\sqrt{I_a^2+I_r^2}$$

This current lags behind the applied voltage in phase by an angle φ whose tangent is

$$\tan\varphi=I_r/I_a$$

The active power in the circuit is the sum of the active powers in the parallel paths

$$P=P_1+P_2=VI_1\cos\varphi_1+VI_2\cos\varphi_2=VI\cos\varphi$$

Similarly, the reactive power in the circuit is

$$Q = Q_1 + Q_2 = VI_1 \sin \varphi_1 + VI_2 \sin \varphi_2 = VI \sin \varphi$$

The apparent power in the circuit is

$$S = \sqrt{P^2 + Q^2}$$

Example 6-3. Given: A circuit having two parallel paths, one containing an inductor of $r_1 = 1$ ohm and $x_{L1} = 3$ ohms, and the other an inductor of $r_2 = 3$ ohms and $x_{L2} = 2$ ohms, both energized by 230 V.

To find: The currents in the paths and the circuit (total) current.

Solution.

$$I_1 = V/z_1 = 230/\sqrt{1^2 + 3^2} = 72.8 \text{ A}$$

$$I_2 = V/z_2 = 230/\sqrt{3^2 + 2^2} = 64 \text{ A}$$

$$\cos \varphi_1 = r_1/z_1 = 1/3.16 = 0.317$$

$$\sin \varphi_1 = x_{L1}/z_1 = 3/3.16 = 0.95$$

$$\cos \varphi_2 = r_2/z_2 = 3/3.6 = 0.833$$

$$\sin \varphi_2 = x_{L2}/z_2 = 2/3.6 = 0.556$$

The component currents in the first parallel path are

$$I_{a1} = I_1 \cos \varphi_1 = 72.8 \times 0.317 = 23 \text{ A}$$

$$I_{r1} = I_1 \sin \varphi_1 = 72.8 \times 0.95 = 69 \text{ A}$$

The component currents in the second parallel path are

$$I_{a2} = I_2 \cos \varphi_2 = 64 \times 0.833 = 53.2 \text{ A}$$

$$I_{r2} = I_2 \sin \varphi_2 = 64 \times 0.556 = 35.4 \text{ A}$$

The components of the circuit (total) current are

$$I_a = I_{a1} + I_{a2} = 23 + 53.2 = 76.2 \text{ A}$$

$$I_r = I_{r1} + I_{r2} = 69 + 35.4 = 104.4 \text{ A}$$

The circuit (total) current is

$$I = \sqrt{I_a^2 + I_r^2} = \sqrt{76.2^2 + 104.4^2} = 129.4 \text{ A}$$

6-7. A Circuit Containing Only a Capacitance

(a) Voltage and Current

If the voltage applied to a capacitor (Fig. 6-17) is

$$v = V_m \sin \omega t$$

then the charge across the capacitor

$$q = Cv = CV_m \sin \omega t$$

varies with the applied voltage (Fig. 6-18).

The current in a circuit containing only a capacitor is equal to the time rate of change of the charge and proportional to the time rate of change of the capacitor voltage

$$i = dq/dt = C\,dv/dt$$

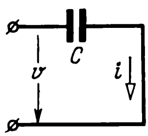


Fig. 6-17. Circuit containing only a capacitance

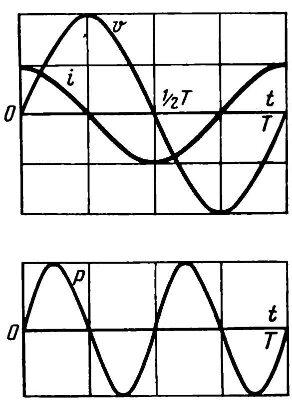


Fig. 6-18. Waveforms of current, voltage and power for a circuit containing only a capacitance

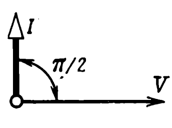


Fig. 6-19. Vector diagram for a circuit containing only a capacitance

When a sinusoidal voltage makes a zero crossing (passes through zero value, Fig. 6-18), its rate of change is a maximum and, as a consequence, so is the current in the circuit containing only a capacitance. At the instants when the voltage reaches its amplitude (peak) value, its rate of change drops to zero, and so does the current in the circuit.

Thus, the current in a circuit containing only a capacitance

$$i = C\,dv/dt = CV_m\,(d \sin \omega t/dt) = C\omega V_m \cos \omega t$$
$$= I_m \sin (\omega t + \pi/2) \tag{6-24}$$

varies sinusoidally and leads the voltage by 90° (it is said to be in quadrature leading with the voltage (see Fig. 6-19).

(b) Capacitive Reactance

It follows from Eq. (6-24) that the amplitude of the current in the circuit of Fig. 6-17 is

$$I_m = C\omega V_m$$

Dividing the above expression by $\sqrt{2}$ gives

$$I = C\omega V = V/(1/\omega C) = V/x_C \tag{6-25}$$

The above expression states Ohm's law for a circuit containing only a capacitance C in terms of rms voltage and current. The quantity

$$x_C = 1/\omega C = 1/2\pi f C \tag{6-26}$$

is called the *capacitive reactance.*

The capacitive reactance of a circuit is inversely proportional to the capacitance and frequency of the a.c. current. As the frequency varies from $f = 0$ (direct current) to $f = \infty$, the capacitive reactance varies from $x_C = \infty$ to $x_C = 0$.

(c) Power

The instantaneous powei in a circuit containing only a capacitance is

$$p = vi = V_m \sin \omega t I_m \cos \omega t = VI \sin 2\omega t$$

A plot of instantaneous power for such a case is shown in Fig. 6-18. As follows from the above equation, the instan-

taneous power in a circuit containing only a capacitance varies at twice the supply frequency, alternatively reaching a positive and a negative peak, $VI = I^2/\omega C$. As the voltage is raised (during the first and third quarters of a cycle, Fig. 6-18), the energy stored by the electric field rises from zero to a maximum value

$$W_m = CV_m^2/2 = CV^2 \tag{6-27}$$

This energy is supplied by an external generator, which means that the circuit is operating as a load and corresponds to the positive value of power.

As the voltage is brought down (during the second and fourth quarters of a cycle, Fig. 6-18), the energy stored by the electric field is decreased from a maximum value to zero, which means that the circuit returns it to the generator. Thus, during these parts of a cycle the circuit is operating as a source of power, which corresponds to the negative value of power. The energy that the circuit receives over a half-cycle is zero, so the average power in the circuit is likewise equal to zero.

The maximum power in a circuit containing only a capacitance is referred to as *reactive power*

$$Q = VI = V^2\omega C = W_m\omega$$

It characterizes the exchange of energy between the circuit and the associated generator.

Example 6-4. Given: A 80-μF capacitor connected in a 380-V, 50-Hz supply line. To find: x_C, I and W_m.

Solution.

$$x_C = 1/2\pi fC = 1/(2\pi \times 50 \times 80 \times 10^{-6})$$

$$= \text{(approx.) } 10^6/25{,}000 = 40 \text{ ohms}$$

$$I = V/x_C = 380/40 = 9.5 \text{ A}$$

$$W_m = CV^2 = 80 \times 10^{-6} \times 380^2 = 11.5 \text{ J}$$

6-8. The Oscillatory Circuit

Initially, when a charged capacitor is connected to an inductor having a negligible resistance (Fig. 6-20), the capacitor voltage is a maximum, V_{Cm}, and the electric field established by the capacitor stores a maximum energy, $W_{Cm} = CV_{Cm}^2/2$.

When the switch is closed, the capacitor begins to discharge and gives rise to a flow of current round the circuit. As a result, the capacitor voltage decreases, and the potential energy of the electric field established by the capacitor is transferred to the kinetic energy of the magnetic field set up around the inductor.

As the capacitor discharges more and more, the current around the circuit gradually rises until it reaches a maximum value—at that instant the capacitor will be completely discharged. The energy that the magnetic field has stored by that time will be

$$W_m = LI_m^2/2 = CV_{Cm}^2/2 \tag{6-28}$$

Although the capacitor voltage falls to zero, the flow of current does not cease in the circuit, because there exists a magnetic field. The current in the circuit is maintained by the emf of self-induction which is positive when the current decreases. The current having the previous direction in the circuit of the discharged capacitor causes electrons to migrate from the formerly negative plate to the formerly positive plate. As a result, the former begins to charge positively and the latter, negatively. In the absence of a resistance in the circuit, this process would go on until the capacitor had charged to a voltage equal in magnitude but opposite in sign to the original one. Then the capacitor would discharge in the reverse direction, the discharge would again be followed by a re-charge, and these events would repeat themselves periodically.

In such a circuit, the energy stored by the electric field is transferred to the magnetic field and back. Obviously, what we have are oscillations, and the circuit is called — quite appropriately — the *oscillatory circuit.* The current in and the voltage across an oscillatory circuit vary sinusoidally (Fig. 6-21), with the capacitor voltage, equal to the

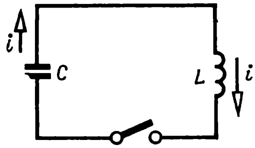


Fig. 6-20. Oscillatory circuit

emf of self-induction at any instant, being shifted by a quarter of a cycle in phase from the current.

The charge on the capacitor varies in proportion to variations in the capacitor voltage

$$dq = i\,dt = C\,dv_C$$

Hence the current in the circuit

$$i = C\,dv_C/dt$$

is seen to be proportional to the time derivative of the capacitor voltage.

The capacitor voltage

$$v_C = -e_L = L\,di/dt \tag{6-29}$$

is proportional to the time derivative of the circuit current. This form of relationship between the two quantities can exist if the current obeys the sine law and the voltage, the cosine law.

To determine the frequency f_0 of the current in such a circuit, we shall first write the current amplitude

$$I_m = V_{Cm}\omega_0 C = V_{Cm} \times 2\pi f_0 C \tag{6-30}$$

On substituting the above equation in Eq. (6-28), we get

$$CV_{Cm}^2/2 = LV_{Cm}^2 \times 4\pi^2 f_0^2 C^2/2$$

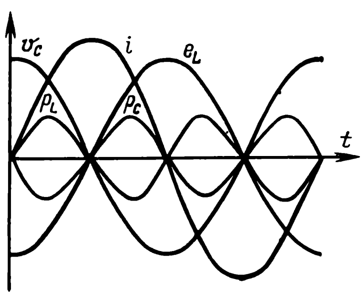


Fig. 6-21. Waveforms of voltage, current and power for an oscillatory circuit

Hence,

$$f_0 = 1/2\pi\sqrt{LC} \tag{6-31}$$
$$\omega_0 = 1/\sqrt{LC}$$

The quantity f_0 is called the *natural frequency* of an oscillatory circuit, and ω_0 is its *natural angular frequency.*

On the basis of Eqs. (6-30) and (6-31), the amplitude of the current in an oscillatory circuit is

$$I_m = \sqrt{CV_{Cm}^2/L} = V_{Cm}/\sqrt{L/C} = V_{Cm}/z_c$$

Here, the quantity $z_c = \sqrt{L/C}$ has the dimension of impedance and is called the *characteristic impedance of the oscillatory circuit.*

If an oscillatory circuit contains a resistance r which does not exceed twice the value of the characteristic impedance, each cycle of oscillation turns some of the energy to heat, so the amplitude of the current and voltage is decreased—the oscillations are said to be damped. If r exceeds the above value, the capacitor will discharge *aperiodically*, with the capacitor voltage and charge reaching zero gradually and continually, that is, with oscillations.

In order to produce *undamped oscillations* in a circuit made up of a capacitance C, an inductance L and a resistance r, one needs an a.c. supply source which would make up for the energy dissipated in the circuit resistance. The source can be connected to the oscillatory circuit in series or parallel. Accordingly, we shall then have a series resonant circuit (Sec. 6-9) and a parallel resonant circuit (Sec. 6-10).

6-9. Voltage (Series) Resonance

When a circuit containing a resistance r, an inductance L and a capacitance C* (Fig. 6-22) carries a sinusoidal current

$$i = I_m \sin \omega t$$

the voltage across such a circuit is the sum of three components (Fig. 6-23), namely an active voltage $V_a = Ir$, which is in phase with the current, an inductive voltage $V_L = Ix_L$, in quadrature leading with the current, and a capacitive

* Usually symbolized as an LCR circuit in the UK and US literature of the subject.— *Translator's note.*

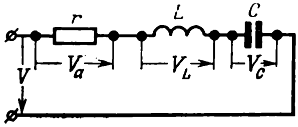


Fig. 6-22. Circuit containing a resistance, an inductance and a capacitance (an LCR circuit)

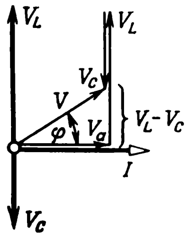


Fig. 6-23. Vector diagram for an LCR circuit at $x_L > x_C$

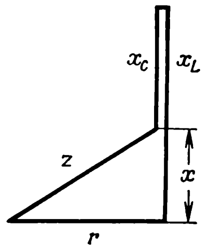


Fig. 6-24. Impedance triangle for an LCR circuit at $x_L > x_C$

voltage $V_C = Ix_C$, in quadrature lagging with the current. The terminal voltage can be found from a right-angled triangle (see Fig. 6-23) one leg of which is the active voltage vector and the other leg is the difference between the inductive and capacitive voltage vectors, or mathematically

$$V = \sqrt{V_a^2 + (V_L - V_C)^2} \tag{6-32}$$

On replacing V_a, V_L and V_C in Eq. (6.32) by the respective products of current, resistance, inductive and/or capacitive reactance, we obtain

$$V = \sqrt{(Ir)^2 + (Ix_L - Ix_C)^2} = I\sqrt{r^2 + (x_L - x_C)^2} = Iz \tag{6-33}$$

Hence, Ohm's law in terms of rms values takes the form

$$I = V/z \tag{6-34}$$

The impedance

$$z = \sqrt{r^2 + (x_L - x_C)^2} = \sqrt{r^2 + x^2} \tag{6-35}$$

may be depicted as the hypotenuze of a right-angled impedance triangle (Fig. 6-24) which can be derived from the voltage triangle, if we divide its sides by the current I.

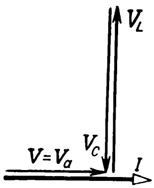


Fig. 6-25. Vector diagram for a series (voltage) resonance

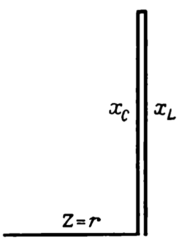


Fig. 6-26. Impedance triangle for an LCR circuit at $x_L = x_C$

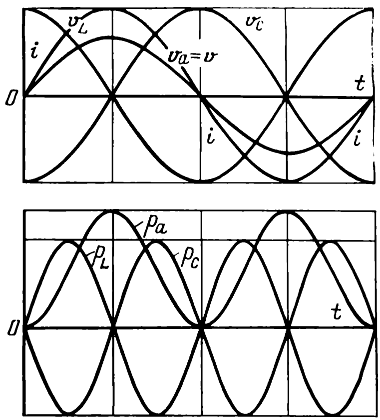


Fig. 6-27. Waveforms of current, voltage and power at series (voltage) resonance

The term $x = x_L - x_C$ is called the *reactive impedance of the circuit.*

The current is shifted from the voltage in phase by an angle φ whose tangent is

$$\tan \varphi = (V_L - V_C)/V_a = (x_L - x_C)/r$$

When $x_L > x_C$ and, as a consequence, $V_L > V_C$ (Figs. 6-23 and 6-24), the current is lagging behind the voltage by an angle φ; when $x_L < x_C$ and $V_L < V_C$, the current is leading the voltage in phase.

When $x_L = x_C$ and, naturally, $V_L = V_C$ (Figs. 6-25 and 6-26), what is known as voltage (or series) resonance takes place. In this condition, the impedance of an oscillatory

circuit is purely resistive

$$z = \sqrt{r^2 + (x_L - x_C)^2} = r$$

In the circumstances, the impedance is a minimum ($z = r$), and at a given terminal voltage V, its current is a maximum

$$I_r = V/r \tag{6-36}$$

Thus, *at resonance the reactances,* x_L *and* x_C, *cancel each other and the current in the circuit is in phase with the voltage across it:*

$$\tan \varphi = x/r = 0, \quad \varphi = 0$$

The inductive voltage V_L and the capacitive voltage V_C, equal in magnitude but opposite in phase (see Figs. 6-25 and 6-27), cancel each other, and the voltage across the circuit is equal to its active, or resistive, voltage.

The ratio of any of the two reactive voltages across an oscillatory circuit to its total voltage is termed the Q (quality) factor of that circuit

$$\begin{aligned} Q &= V_L/V = V_C/V = I_r x_L/I_r r = I_r x_C/I_r r \\ &= x_L/r = x_C/r = z_c/r \end{aligned} \tag{6-37}$$

The Q-factor shows how many times V_L and V_C exceed the terminal voltage V of an oscillatory (or resonant) circuit at resonance. At high values of Q, V_L and V_C are many times the terminal voltage V.

When V_L and V_C are equal and the phase difference between them is a half-cycle, this is an indication that at any instant of time the capacitor and inductor voltages are equal in magnitude but opposite in sign, $v_L = -v_C$. As a consequence, the instantaneous inductive and capacitive powers at any time are equal in magnitude but opposite in sign, $p_L = -p_C$ (Fig. 6-27), because $p_L = iv_L$, and $p_C = iv_C$.

Hence, we may state that an increase in the energy stored by the magnetic field occurs solely owing to a decrease in the energy stored by the electric field and vice versa. Thus, the only thing left for the generator to do is to make up for the energy dissipated in the resistance.

To sum up, voltage (or series) resonance is characterized by the fact that the magnetic and electric fields set up

by an oscillatory circuit periodically exchange all of the energy they respectively store.

At voltage (series) resonance,

$$\omega L = 1/\omega C \quad \text{or} \quad \omega^2 LC = 1 \tag{6-38}$$

Hence, the angular resonant frequency is

$$\omega = 1/\sqrt{LC} = \omega_0 \tag{6-39}$$

and the resonant frequency is

$$f = \omega/2\pi = 1/2\pi\sqrt{LC} = f^0 \tag{6-40}$$

In other words, resonance takes place at a generator frequency equal to the natural frequency of the oscillatory (resonant) circuit.

When one adjusts the circuit parameters in order to secure resonance, one is said to tune the circuit to resonance.

At resonance, the parameters ω, L and C are related by Eq. (6-38) from which it follows that a circuit can be tuned to resonance in one of several ways. For example, if we hold ω and L constant, this can be done by adjusting C. If we hold L and C constant, this can be done by adjusting the frequency ω of the supply generator. If we hold ω constant, this can be done by adjusting L and C, and so on.

Figure 6-28 shows plots of reactances, namely $x_L = \omega L$, $x_C = 1/\omega C$ and $x = x_L - x_C$, as functions of frequency $\omega = 2\pi f$. They are called the frequency responses of a series resonant circuit.

The inductive reactance $x_L = \omega L$ increases in proportion to frequency ω from 0 at $\omega = 0$ to infinity at $\omega = \infty$. The capacitive reactance $x_C = 1/\omega C$ varies inversely proportional to frequency from $-\infty$ to zero. The reactance $x = x_L - x_C$ varies from $x = -\infty$ to $x = 0$ and then to $x = \infty$ as the frequency varies from $\omega = 0$ to the resonant frequency $\omega = \omega_0$, and then to $\omega = \infty$.

If an LCR circuit is held at a constant voltage V while the frequency ω is varied, all the quantities defining the operating conditions in the circuit will also vary. Among other things, the circuit current $I = V/z = V/\sqrt{r^2 + x^2}$ will be zero at $\omega = 0$ and $\omega = \infty$, and a maximum, $I = V/r$, at the resonant frequency $\omega = \omega_0$ (Fig. 6-29).

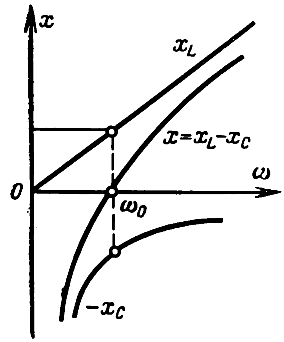


Fig. 6-28. Frequency response curves

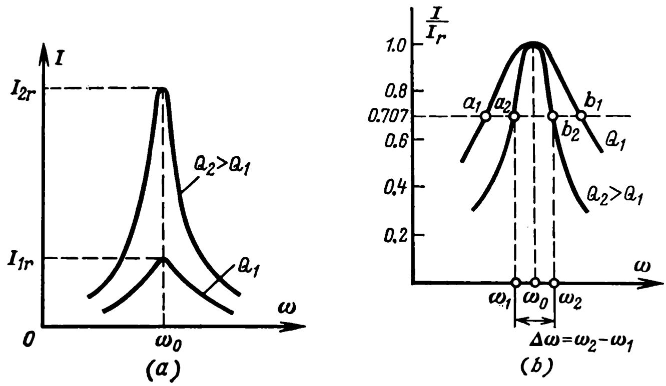


Fig. 6-29. Resonance curves for different values of the Q-factor

The so-called resonance curves, $I = f(\omega)$, of a series resonant circuit for two values of the Q-factor, Q_1 and $Q_2 > Q_1$, and for the same values of V, L and C are shown in Fig. 6-29a. Similar curves are shown in Fig. 6-29b, but here the ordinates are the current values relative to the value at resonance, that is, I/I_r, and not the absolute values, so the resonance curve is $I/I_r = f(\omega)$.

As is seen from the resonance curves, current oscillations are at their maximum only on frequencies close to the natural frequency ω_0 of the circuit. This is another way of saying that the circuit sustains oscillations within a particular frequency band or range only, called the bandwidth of the

circuit. It is customary to define the bandwidth of a circuit as that within which the circuit current is not less than $I_r/\sqrt{2} = 0.707\ I_r$.

If we draw a straight line parallel to the x-axis (Fig. 6-29b) at an ordinate of 0.707 and drop perpendiculars from points a and b where the line cuts the resonance curve, onto the x-axis, the intercepts thus obtained will mark the frequencies, ω_1 and ω_2, that limit the bandwidth

$$\Delta\omega = \omega_2 - \omega_1$$

Quite appropriately, the two frequencies are called limiting.

From Fig. 6-29 it follows that the higher values of the Q-factor correspond to the narrower resonance curves and the narrower bandwidths, $\Delta\omega$.

The property of resonance is widely utilized in many applications, notably in radio and electronics. In power circuits, however, voltage (series) resonance is an abnormal condition and may have dangerous consequences because of an excessive increase in current and overvoltages across the reactive elements of the circuit.

6-10. Current (Parallel) Resonance

(a) A Lossless Parallel Resonant Circuit

Consider a parallel circuit (Fig. 6-30) one branch of which contains inductance L and the other a capacitance C. If the two branches have the same reactance

$$\omega L = 1/\omega C$$

the circuit will be in a condition known as *current* (or *parallel*) *resonance.*

From the above equation it follows that the condition of resonance can be secured by adjusting the inductance, capacitance, or frequency, because

$$L = 1/\omega^2 C; \quad C = 1/\omega^2 L \quad \text{and} \quad \omega = 1/\sqrt{LC} = \omega_0 \qquad (6\text{-}41)$$

At current resonance, the branch currents

$$I_1 = I_L = V/\omega L = I_2 = I_C = V\omega C$$

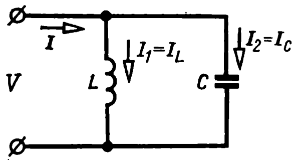


Fig. 6-30. Parallel LC circuit

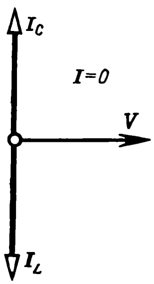


Fig. 6-31. Vector diagram of parallel (current) resonance at $r = 0$

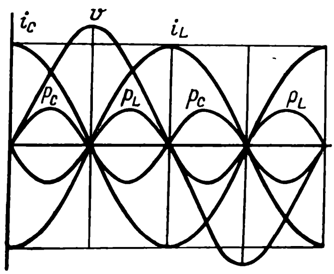


Fig. 6-32. Waveforms of current, voltage and power at parallel (current) resonance ($r = 0$)

are equal in magnitude and vary in anti-phase (Fig. 6-31), because I_L is in quadrature lagging and I_C is in quadrature leading with the voltage.

By Kirchhoff's current (first) law, the current in a series circuit (the total or circuit current) is

$$i = i_L + i_C$$

Since, however, $i_C = -i_L$, then $i = i_L + i_C = 0$, or in words, the total (circuit) current is zero.

Plots of currents, voltage and power applying to this case are given in Fig. 6-32.

The fact that the circuit contains no resistance indicates that the energy stored by the circuit is not dissipated.

During the first quarter of a cycle (see Fig. 6-32), the capacitor voltage rises from zero to a maximum value V_{Cm}, and the electric field it sets up stores the energy $W_{Cm} = CV_{Cm}^2/2$. During the next quarter of a cycle, the electric field collapses and gives up the energy it has stored.

During the same first quarter of a cycle, the current in the inductor decreases from a maximum I_{Lm} to zero, the

associated magnetic field collapses, and it gives up the energy it has stored. During the next quarter of a cycle, the current in the inductor rises to I_{Lm}, and the energy stored by the magnetic field of the inductor rises from zero to a maximum $W_{Lm} = LI^2_{Lm}/2$.

From the foregoing and Fig. 6-32, it is an easy matter to see that during the first quarter of a cycle the kinetic energy of the magnetic field is converted to the potential energy of the electric field; during the second quarter of a cycle the sequence is reversed, namely the energy of the electric field is converted to that of the magnetic field. This sequence of energy exchange between the electric and magnetic fields is repeated periodically.

There is no exchange of energy between the circuit and its power source, because the current in the series (common) part of the circuit is zero.

(b) A Lossy Parallel Resonant Circuit

The circuit of Fig. 6-33 consists of an inductor and a capacitor connected in parallel and held at the same voltage, V.

The inductor current is

$$I_1 = V/z_1 = V/\sqrt{r_1^2 + x_{L1}^2}$$

This current lags in phase behind the voltage by an angle whose tangent is

$$\tan \varphi_1 = x_{L1}/r_1$$

The inductor current can be resolved into two components, namely an active current, $I_{a1} = I_1 \cos \varphi_1$, which is in phase with the applied voltage, and a reactive current, $I_{r1} = I_L = I \sin \varphi_1$, which is in quadrature lagging with the applied voltage (Fig. 6-34).

The capacitor current is

$$I_2 = I_C = V/x_C = V/(1/\omega C) = V\omega C$$

It is in quadrature leading with the applied voltage.

The total (circuit) current can be found from the current triangle of Fig. 6-34, where one leg is the active current, $I_a = I_{a1}$, and the other leg is the reactive current equal to the difference between the reactive component of the

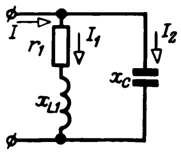


Fig. 6-33. Parallel LCR circuit

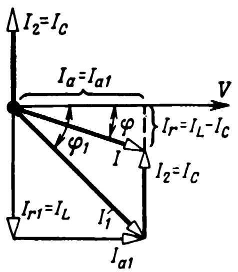


Fig. 6-34. Vector diagram for a parallel LCR circuit

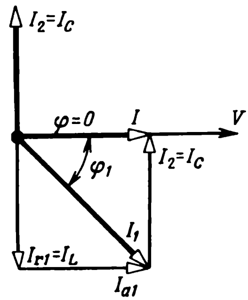


Fig. 6-35. Vector diagram at current (parallel) resonance

inductor current and the capacitor current

$$I_r = I_{r1} - I_2 = I_L - I_C$$

Hence, the total (circuit) current is

$$I = \sqrt{I_a^2 + I_r^2}$$

The phase difference between the total (circuit) current and the applied voltage can be deduced from its tangent (Fig. 6-34)

$$\tan \varphi = I_r/I_a = (I_L - I_C)/I_{a1}$$

The total current may lag behind the voltage by an angle φ when $I_L > I_C$ or lead it when $I_L < I_C$, or be in phase with it (Fig. 6-35) when $I_L = I_C$. In the last case, the circuit is in a condition called current (or parallel) resonance when $I = \sqrt{I_a^2 + I_r^2} = I_a$, and the power $P = VI \cos \varphi = VI$, because $\varphi = 0$, and $\cos \varphi = 1$.

To sum up, the total (circuit) current is equal to the active current of the inductor and is less than the total inductor current ($I_{a1} < I_1$).

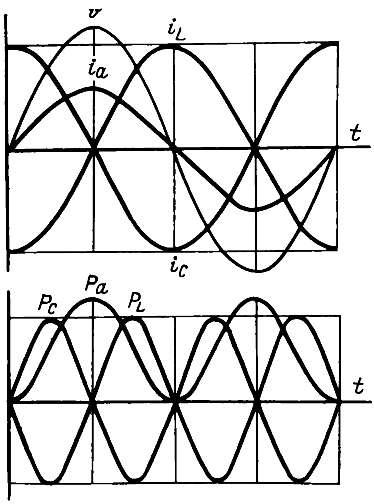


Fig. 6-36. Waveforms of current, voltage and power for a parallel LCR circuit at $I_C = I_L$

The ratio of the capacitor or inductor current ($I_1 \approx I_2$) to the total (circuit) current at resonance is called the Q (quality) factor of a parallel resonant circuit:

$$I_1/I = Q$$

It shows how many times the current in a parallel resonant circuit exceeds the total (circuit) current at resonance.

In this case the maximum power expended to set up the magnetic field VI_L is equal to the maximum power expended to establish the electric field VI_C. Accordingly, the maximum energy stored by the magnetic field is equal to that stored by the electric field, $W_{Lm} = W_{Cm}$. As is the case with the resonant circuit examined earlier, the energy stored by the magnetic field is completely transferred to the electric field during one quarter of a cycle and back during the next. The generator supplies only that part of the total energy which is dissipated in the circuit resistance. Because the reactive components of current cancel each other, the generator circuit solely carries the active current associated with the energy lost in the resistance. Plots of the currents, voltages and power applicable to the resonant circuit of Fig. 6-33 appear in Fig. 6-36.

6-11. The Power Factor

The rated current, I_n, of an electric machine, transformer or electromagnetic apparatus is specified on the basis of the maximum allowable temperature rise of its windings, and its rated voltage V_n, on the basis of the maximum allowable temperature rise of the magnetic circuit. Thus, the rated current and the rated voltage are the maximum allowable values under service conditions. Accordingly, a generator can develop a maximum active power when its power factor is $\cos \varphi = 1$, because

$$P_n = V_n I_n \cos \varphi = V_n I_n = S_n \qquad (6\text{-}42)$$

In this case, the generator is utilized most effectively, since its active power is equal to the rated apparent power, S_n.

As the power factor decreases, the active power developed by the generator also decreases in proportion to $\cos \varphi$. For example, at $\cos \varphi = 0.5$, the active power of a generator

$$P_n = V_n I_n \times 0.5 = 0.5 S_n$$

is only half its rated apparent power. In other words, the rated power of the generator is utilized incompletely, which is of course a waste of input energy.

The power factor of a generator (or a generating station) depends on the load(s) it carries. This is why power authorities require that each consumer should maintain a power factor of at least 0.9 to 0.92. If a consumer fails to do so, he is charged an extra payment. If, on the other hand, a consumer maintains a power factor at a higher value, his electricity bill is discounted.

In industry, most loads are electric motors and other electromagnetic equipment, which depend for their operation on an alternating magnetic flux.

Such a magnetic flux can be produced only if the circuit current has a reactive component $I_{r1} = I_L$ (see Figs. 6-33 and 6-34), because it controls the reactive power of a load, $Q = VI_r$.

Electricity cannot be converted to any other form of energy (mechanical, thermal, etc.) unless the circuit current has an active component $I_{a1} = I_a$ (see Figs. 6-33 and 6-34),

as it controls the active power in the load, $P = VI_a$. The current in a load, such as an induction motor, can be expressed (see Eq. (6-23a) and Fig. 6-34), as follows

$$I_1 = \sqrt{I_{a1}^2 + I_{r1}^2}$$

This current, I_1, lags behind the applied voltage by an angle φ_1 for which $\tan \varphi_1 = I_{r1}/I_{a1}$ and the power factor is $\cos \varphi_1 = I_{a1}/I_1$.

The current in a motor can be found, using a well-known power equation

$$I_1 = P_1/V \cos \varphi_1 = (P_1/V)\,(1/\cos \varphi_1)$$

It follows from the above equation that when P_1 and V are held constant, a decrease in $\cos \varphi_1$ will cause I_1 to increase in inverse proportion to $\cos \varphi_1$, and the power loss in the wires connecting the motor and generator (station) will vary in proportion to the current squared ($\Delta P = I_1^2 r_w$), that is, in inverse proportion to the power factor squared

$$\Delta P = I_1^2 r_w = [(P_1/V)\,(1/\cos \varphi_1)]^2\, r_w$$
$$= (P_1^2/V^2)\, r_w\, (1/\cos^2 \varphi_1)$$

This is a second cause why the power factor of a load should be maintained at 0.9 to 0.92.

At no-load, the power factor of a motor is $\cos \varphi_{\text{no-load}} = 0.1$ to 0.3, and at rated load it is $\cos \varphi_n = 0.8$ to 0.85.

The power factor can be improved:

(1) by increasing load on motors and by maintaining it close to the nominal (rated) value;[1]

(2) by replacing the underloaded motors by those of a lower rating so that they can operate at a load close to the rated one;

(3) by providing a synchronous motor which, when sufficiently excited, will supply a leading reactive current to the circuit, (see Sec. 10-14);

(4) by connecting a bank of capacitors in parallel with the motor.

If we connect a bank of capacitors in parallel with the load, such as an induction motor operating at $\cos \varphi_1$ (see Fig. 6-33), a capacitive current $I_C = I_2$ will begin to flow in the circuit in quadrature leading with the applied voltage.

The current I in the line wires, equal to the sum of I_1 and I_2, will be less than I_1 and make a phase angle $\varphi < \varphi_1$ with the voltage (see Fig. 6-34). In this way the power factor of the plant will be improved because all (or part) of the reactive current in the motor I_{r1} will be compensated for by the capacitor current $I_C = I_2$.

In such a case, an inductive load (motor) will store energy in its magnetic field during, say, the first and third quarters of a cycle, and give it up during the second and fourth. In contrast, the capacitor will give up the stored energy during the first and third quarters of a cycle, and accumulate it during the second and fourth. To sum up, the magnetic field of the motor stores all or part of the energy supplied by the electric field due to the capacitor, while the generator and the transmission line between the generator and motor are partly or completely relieved of the energy being exchanged.

When a bank of capacitors is connected to the circuit, the total (circuit) current is

$$I = \sqrt{I_a^2 + I_r^2}$$

and the resultant reactive current of the installation is

$$I_r = I_{r1} - I_{r2}$$

The tangent of the phase difference for the motor is

$$\tan \varphi_1 = I_L / I_a$$

and the reactive current in the motor is

$$I_L = I_a \tan \varphi_1$$

The tangent of the phase difference φ for the motor-capacitor combination (see Fig. 6-34) is

$$\tan \varphi = (I_a \tan \varphi_1 - I_C)/I_a$$

Hence,

$$I_C = I_a \tan \varphi_1 - I_a \tan \varphi = I_a (\tan \varphi_1 - \tan \varphi)$$
$$= (P/V) (\tan \varphi_1 - \tan \varphi)$$

Since $I_C = V\omega C$
then

$$V\omega C = (P/V) (\tan \varphi_1 - \tan \varphi)$$

Using the above equation, it is an easy matter to determine the capacitance for a bank of capacitors, required to improve the power factor of an installation

$$C = (P/\omega V^2)(\tan \varphi_1 - \tan \varphi)$$

In most cases, the reactive current is compensated for incompletely, and one is usually content with having $\cos \varphi$ as high as 0.95. Any further improvement in the power factor would require a bigger bank of capacitors, which would be unattractive economically.

Power-factor improvement is a pivotal problem in electrical engineering. A higher value of $\cos \varphi$ means a saving in electricity, because this would cut any forms of power loss and improve the utilization of the installed capacity at the generating and other levels of a power system.

Example 6-5. Given: A load with a voltage of $V = 400$ V and a power $P = 50$ kW. The resistance of the wires connecting load and generator is $r = 0.04$ ohm.

To find: The power lost in the wires at $\cos \varphi_1 = 0.9$ and $\cos \varphi_2 = 0.3$.

Solution.

Under the specified conditions, the load and wire currents are

$$I_1 = P/(V \cos \varphi_1) = 50{,}000/(400 \times 0.9) = 139 \text{ A}$$
$$I_2 = P/(V \cos \varphi_2) = 50{,}000/(400 \times 0.3) = 417 \text{ A}$$

The power lost in the wires is

$$\Delta P_1 = I_1^2 r = 139^2 \times 0.04 = 19{,}320 \times 0.04 = 772 \text{ W}$$
$$\Delta P_2 = I_2^2 r = 417^2 \times 0.04 = 173{,}900 \times 0.04 = 6956 \text{ W}$$

The same loss of power in per cent of the power in the load is

$$\Delta P_1 = 772 \times 100 \div 50{,}000 = 1.54\%$$
$$\Delta P_2 = 6956 \times 100 \div 50{,}000 = 13.9\%$$

Example 6-6. Given: A motor operating on $V = 220$ V, 50 Hz, $\cos \varphi = 0.6$, and developing $P = 20$ kW.

To find: The capacitance for the bank of capacitors to be installed in parallel with the motor so as to raise $\cos\varphi$ to 0.9.

Solution.

Using the equation derived earlier, the capacitance of the capacitor bank is found to be

$$C = (P/\omega V^2)(\tan\varphi_1 - \tan\varphi)$$
$$= [20{,}000/(2 \times 3.14 \times 50 \times 220^2)](1.327 - 0.484)$$
$$= 0.00113\ \text{F} = 1130\ \mu\text{F}$$

Thus, the bank of capacitors to be connected in parallel with the motor must have a capacitance of 1130 μF.

6-12. Active and Reactive Energy

The electric energy expended in an a.c. circuit over a time t is called *active energy*. If the power is held constant, the active energy is given by

$$W_a = Pt = VI\cos\varphi t \qquad (6\text{-}43)$$

If the power is varying, the active energy is found as a sum

$$P_1t_1 + P_2t_2 + \ldots = W_{a1} + W_{a2} + \ldots = W_a$$

In the above expression, each term is the energy that the circuit receives over a time interval t_1, t_2, t_3, etc., within which the power remains unvarying.

The product of an unvarying reactive power Q and time t is called the *reactive energy*

$$W_r = Qt = VI\sin\varphi t \qquad (6\text{-}44)$$

If the reactive power is varying, the reactive energy is found as a sum

$$Q_1t_1 + Q_2t_2 + \ldots = W_{r1} + W_{r2} + \ldots = W_r$$

If the active and reactive powers are unvarying, the ratio

$$W_a/\sqrt{W_a^2 + W_r^2} = \frac{VI\cos\varphi t}{\sqrt{(VI\cos\varphi t)^2 + (VI\sin\varphi t)^2}}$$
$$= \frac{VIt\cos\varphi}{VIt\sqrt{\cos^2\varphi + \sin^2\varphi}} = \cos\varphi$$

is the power factor.

If the active and reactive powers are varying with time, it is usual to measure them over a specified time interval by a watthour-meter and a varhour-meter and to determine the ratio

$$W_a/\sqrt{W_a^2+W_r^2}=\cos\varphi_{av}$$

which is called the *average power factor of an installation.* It is an important indicator of technical and economic performance.

Example 6-7. Given: Watthour-meter readings 2326 kWh and 2476 kWh and varhour-meter readings 1673 kVArh and 1773 kVArh for the start and end of a month.

To find: The average power factor.

Solution.

The active energy expended over the month is

$$W_a = 2476 - 2326 = 150 \text{ kWh}$$

The reactive energy over the same interval is

$$W_r = 1773 - 1673 = 100 \text{ kVArh}$$

Hence, the average power factor is

$$\cos\varphi_{av}=W_a/\sqrt{W_a^2+W_r^2}=150/\sqrt{150^2+100^2}$$
$$=0.83$$

Chapter Seven Three-Phase Networks

7-1. Three-Phase Systems

A *three* (or *poly*)*-phase system* is an assemblage of three (or more) a.c. networks operating at the same frequency, in which the emfs differ in epoch (initial phase) and are generated by a common source.

Three-phase systems have found an extremely wide use because they provide for a more economical power transmission in comparison with single-phase systems. Also, they use relatively simple and reliable generators, motors, and transformers.

M. O. Dolivo-Dobrovolsky (1862-1919)

The credit for the invention of a three-phase system, a three-phase generator, a three-phase motor and a three-phase transformer is due to M. O. Dolivo-Dobrovolsky of Russia.

The individual circuits of a three-phase system are referred to as phases. Together, they make up a *three-phase circuit* or *network*.

The currents, voltages or emfs existing in the phases of a three-phase circuit or network constitute a three-phase system of currents, voltages or emfs.

A simple three-phase generator (Fig. 7-1) is similar in construction to a single-phase machine (see Fig. 5-2), except that its armature carries three identical windings (phase

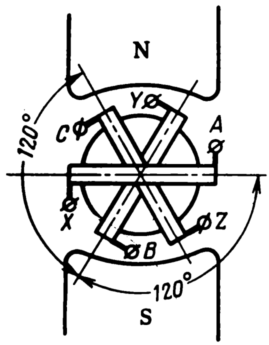


Fig. 7-1. Elementary three-phase generator

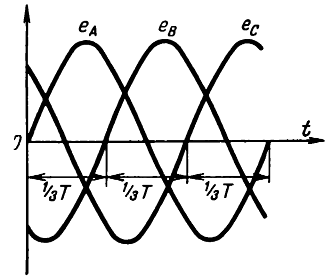


Fig. 7-2. Waveforms of symmetrical emfs in a three-phase system

windings or simply phases), the starts and finishes of which are respectively designated A, B, C and X, Y, Z. The axes of the windings are spaced an equal angle apart. This angle is equal to 120° or $2\pi/3$. Accordingly, the emfs induced in the windings have the same amplitude but are spaced 120° or one-third of a cycle apart. Such three emfs form a *symmetrical system*. If the emfs differ in amplitude or phase angle, they make up an *unsymmetrical system*.

On setting the start of a cycle for the emf in the first phase (phase A) as a reference (or datum) point ($t = 0$), we can write

$$e_A = E_m \sin \omega t \tag{7-1}$$

The emf of the second phase (phase B) will then lag behind that in the first by a third of a cycle, so

$$e_B = E_m \sin (\omega t - 2\pi/3) \tag{7-2}$$

The emf in the third phase (phase C) will then lag behind e_A by two-thirds of a cycle or lead e_A by one third of a cycle, so

$$e_C = E_m \sin (\omega t - 4\pi/3) = E_m \sin (\omega t + 2\pi/3) \tag{7-3}$$

Plots and vector diagrams of the phase emfs are shown in Figs. 7-2 and 7-3.

By convention it is assumed that the positive direction for the emfs in the windings of a three-phase generator is

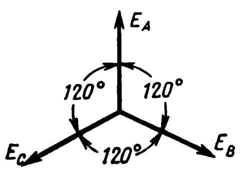


Fig. 7-3. Vector diagram for symmetrical emfs

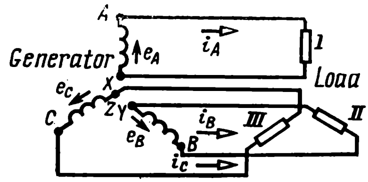


Fig. 7-4. Non-interconnected three-phase system

from the finishes X, Y and Z towards the starts A, B and C of the windings.

If we connected each winding of a three-phase generator to a separate load (Fig. 7-4), we would obtain a non-interconnected six-wire three-phase system. It is uneconomical and so it is never used in practice. As a rule, the windings of a three-phase generator are connected either into a star (or wye, Y), or into a delta (or mesh). With this arrangement, a three-phase system needs only three or four wires instead of six.

For three-phase networks in the USSR, the standard voltages are 127 V, 220 V, 380 V, 660 V, and higher.

7-2. A Generator with Star-Connected Windings

In the case of a star connection, the finishes X, Y and Z of the windings are connected to a common point, called the *neutral point* of a generator (Fig. 7-5). In a four-wire three-phase system, the neutral point is connected to a *neutral wire*. The starts of the phase windings are connected to three line wires.

The voltages between the starts and finishes of the phases or, which is the same, the voltages between each line wire and the neutral wire are known as the *phase voltages*; they are symbolized V_A, V_B and V_C, or, more generally, V_p. If we neglect the voltage drop across the generator windings, the phase voltages may be taken equal to the emfs induced in the respective generator windings.

The voltages between the starts of the windings or, which is the same, between the line wires are called the *line volt-*

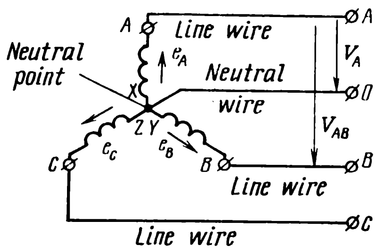


Fig. 7-5. Generator with star-connected windings

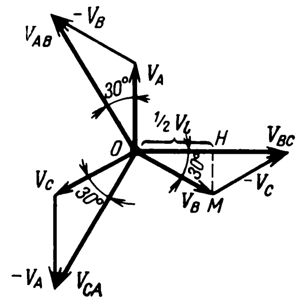


Fig. 7-6. Vector diagram for voltages in a three-phase network

ages; they are symbolized V_{AB}, V_{BC}, and V_{CA}, or more generally, V_l.

Let us establish the relationship between the line and phase voltages for a generator in which the windings are star-connected. Because the finish of the first phase X is connected to the finish Y, rather than the start of the second phase, the instantaneous line voltage between the wires A and B will, according to Kirchhoff's second (voltage) law, be equal to the difference of the respective voltages, namely

$$v_{AB} = v_A - v_B$$

Similarly, the instantaneous values of the other two line voltages will be

$$v_{BC} = v_B - v_C \quad \text{and} \quad v_{CA} = v_C - v_A$$

As is seen, the *instantaneous value of a line voltage is equal to the algebraic difference of the instantaneous values of the respective phase voltages.*

Because v_A, v_B and v_C vary sinusoidally and have the same frequency, the line voltages v_{AB}, v_{BC} and v_{CA} likewise vary sinusoidally, and their rms values can be deduced from the vector diagram of Fig. 7-6 as follows

$$\bar{V}_{AB} = \bar{V}_A - \bar{V}_B$$
$$\bar{V}_{BC} = \bar{V}_B - \bar{V}_C$$
$$\bar{V}_{CA} = \bar{V}_C - \bar{V}_A$$

It follows from the foregoing that the *vector of a line voltage is equal to the difference between the vectors of the respective phase voltages.*

The phase voltages v_A, v_B and v_C have a phase difference of 120° between one another. To determine the line-voltage vector $\bar{V}_{AB}$, we should subtract vectorially the vector $\bar{V}_B$ from the vector $\bar{V}_A$ or, which is the same, add to the latter the vector $\bar{V}_B$ taken with a "minus" sign.

Similarly, the line-voltage vector $\bar{V}_{BC}$ can be found as the difference between the voltage vectors $\bar{V}_B$ and $\bar{V}_C$ and the line-voltage vector $\bar{V}_{CA}$ as the difference between the vectors $\bar{V}_C$ and $\bar{V}_A$.

If we drop a perpendicular from the terminal point of an arbitrarily chosen phase voltage vector, say, $\bar{V}_B$, on the line-voltage vector $\bar{V}_{BC}$, we obtain a right triangle, *OHM*, from which it follows that

$$V_l/2 = V_p \cos 30° = V_p\,(\sqrt{3}/2)$$

so

$$V_l = V_p \sqrt{3} \tag{7-4}$$

Referring to the vector diagram of Fig. 7-6 and Eq. (7-4), we can state that the *rms value of a line voltage is* $\sqrt{3}$ *times the rms value of phase voltage and that the line voltage* V_{AB} *leads the phase voltage* V_A *by* 30°. The same leading phase difference exists between the line voltage V_{BC} and the phase voltage V_B, and also between the line voltage V_{CA} and the phase voltage V_C.

The phase difference between the line voltages is the same as it is between the phase voltages, that is, 120°. But the *star of the line voltage vectors is turned relative to the star of the phase voltages through* 30° *in the positive direction.*

It is to be noted that the above relationships between the line and phase voltages apply only to a symmetrical system of voltages.

Because the line-voltage vectors are found as the difference between the vectors of phase voltages, we can connect the terminal points of the phase-voltage vectors and form a line-voltage vector triangle (Fig. 7-7).

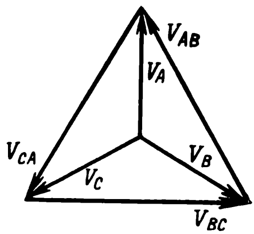


Fig. 7-7. Vector diagram for voltages in a generator with star-connected windings

When the neutral wire is used in operation, the three-wire, three-phase system becomes a four-wire, three-phase system. An advantage of such an arrangement is that one obtains two systems of voltages, namely a system of phase voltages when each load is connected between the neutral wire and any of the line wires, and a system of line voltages when each load is connected between two line wires.

Example 7-1. Given: A generator with a phase voltage of 127 V and 220 V. To find: The line voltage.

Solution.

$$V_l = V_p \sqrt{3} = 127 \times 1.73 = 220 \text{ V}$$

or

$$V_l = V_p \sqrt{3} = 220 \times 1.73 = 380 \text{ V}$$

7-3. A Generator with a Delta-Connected Windings

When the windings of a three-phase generator are connected in a delta (Fig. 7-8), the finish of the first phase, X, is connected to the start of the second phase, B, the finish of the second phase, Y, to the start of the third phase, C, and the finish of the third phase, Z, to the start of the first phase, A. The three line wires from the respective loads are taken to the phase starts A, B and C.

From Fig. 7-8 it is seen that in this arrangement the *phase voltages are the same as the line voltages*

$$V_{AB} = V_A, \quad V_{BC} = V_B, \quad V_{CA} = V_C \tag{7-5}$$

The delta-connected windings of a three-phase generator form a closed mesh which has a very low resistance. Obviously this form of connection may only be used where the emfs

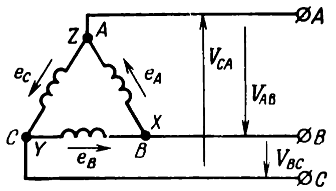


Fig. 7-8. Generator with delta-connected windings

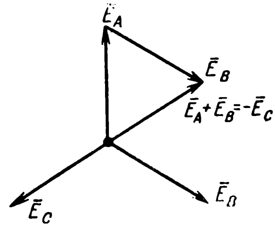


Fig. 7-9. Vector diagram for emfs in a generator with delta-connected windings

existing in this mesh add up to zero. Otherwise, a considerable current would be flowing even in the absence of load, and the generator would be overheated.

The sum of the three symmetrical emfs existing in the generator windings is zero. This can readily be proved, if we add together the emf vectors. Referring to Fig. 7-9, there are three emfs. Combining $\bar{E}_A$ and $\bar{E}_B$ gives a vector equal in magnitude but opposite in direction to the vector $\bar{E}_C$

$$\bar{E}_A + \bar{E}_B = -\bar{E}_C$$

Hence, the sum of the three emf vectors is zero:

$$\bar{E}_A + \bar{E}_B + \bar{E}_C = 0 \qquad (7\text{-}6)$$

A dangerous situation may arise if the generator windings are delta-connected in a wrong way. One of such connections is illustrated in Fig. 7-10, where the finish of the first phase, X, is correctly connected to the start of the second phase, B, but the finish of the second phase, Y, is incorrectly connected to the finish, Z, and not to the start, C, of the third phase, as it ought to have been done. Because of this, the emf $\bar{E}_C$ is subtracted from, rather than added to the sum of the other two emfs. The resultant emf can be found from the vector diagram of Fig. 7-11, where the vectors $\bar{E}_A$, $\bar{E}_B$ and $-\bar{E}_C$ are combined. The sum of these vectors is seen to be equal to twice the vector $\bar{E}_C$:

$$\bar{E}_A + \bar{E}_B - \bar{E}_C = -2\bar{E}_C$$

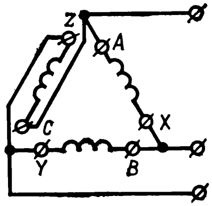


Fig. 7-10. Incorrect delta-connection of generator windings

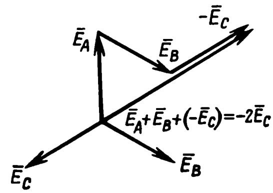


Fig. 7-11. Vector diagram for emfs in the generator with windings connected as shown in Fig. 7-10

Thus, the emf of the closed mesh is now equal in absolute value to twice the phase emf. Recalling that the closed mesh (the generator windings) has a low resistance, this connection is equivalent to a short circuit.

7-4. Star-Connected Loads

Like the generator windings, loads can be star-connected, and the system thus obtained may be either a *four-wire type*, such as in the case of lighting load, or a *three-wire type*, as in power circuits.

In a four-wire, three-phase system, the lamps are connected between the neutral wire and each line conductor (Fig. 7-12) and the rated voltage of the lamps must be equal to the phase voltage of the supply system (mains). With this arrangement, the loads operate under the same conditions as in a single-phase system, because the neutral wire provides for the equality between the phase voltages of the generator and the respective phase voltages of the loads.

As is seen from Fig. 7-12, the currents in the line wires are equal to the currents in the respective phases of the load and generator:

$$I_l = I_p \tag{7-7}$$

The phase currents in the load can be found in the same way as for a single-phase system, namely

$$I_A = V_A/z_A, \quad I_B = V_B/z_B, \quad I_C = V_C/z_C$$

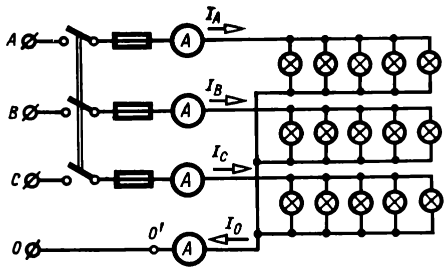


Fig. 7-12. Star-connected loads with a neutral wire

The phase difference between the currents and the respective phase voltages can be determined from their cosines

$$\cos \varphi_A = r_A/z_A, \quad \cos \varphi_B = r_B/z_B, \quad \cos \varphi_C = r_C/z_C$$

or from their tangents

$$\tan \varphi_A = x_A/r_A, \quad \tan \varphi_B = x_B/r_B, \quad \tan \varphi_C = x_C/r_C$$

By Kirchhoff's first (current) law, the instantaneous current in the neutral wire is given by

$$i_0 = i_A + i_B + i_C$$

The rms current in the neutral wire can be found by combining the phase currents vectorially, that is,

$$\bar{I}_0 = \bar{I}_A + \bar{I}_B + \bar{I}_C \tag{7-8}$$

Example 7-2. Given: Generator phase voltage $V_p = 125$ V, load phase impedance $z_A = z_B = r_A = r_B = 12.5$ ohms, $z_C = r_C = 25$ ohms. To find: The phase currents.

Solution.

$$I_A = I_B = 125 \div 12.5 = 10 \text{ A}$$

$$I_C = 125 \div 25 = 5 \text{ A}$$

A vector diagram of the phase voltages and currents is shown in Fig. 7-13. Vectorial combination of the phase currents gives the current in the neutral wire to be

$$I_0 = 5 \text{ A}$$

As is seen, it lags behind V_A by $\varphi_0 = 60°$ in phase.

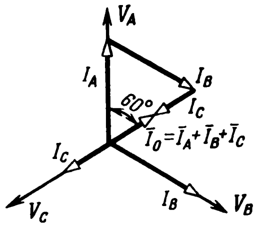


Fig. 7-13. Vector diagram for a four-wire, three-phase network operating into a resistive load

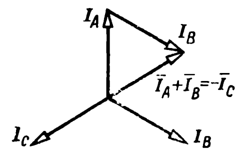


Fig. 7-14. Symmetrical currents in a three-phase network

The cross-sectional area of the neutral wire may be equal to that of the line conductors or be a half or a third as large, because it has to carry a smaller current than the line conductors.

It is to be stressed once more that, whatever the phase load, the neutral wire provides for the equality between the phase voltages of the loads. Should the neutral wire be broken and the load phases differ in impedance, the phase voltages at the load will be likewise different, being lower where the phases have a lower impedance, and higher where the phases have a higher impedance. This is an abnormal condition, which is especially dangerous if a break in the neutral wire is accompanied by a short-circuit in a phase conductor. Then the voltage in the other phases will rise $\sqrt{3}$ times, and all the lamps connected in those phases will blow. This is why, in order to avoid an open-circuit in the neutral wire no fuses or switches are installed in it.

If load on the three phases is the same (such as in the case of a three-phase motor), the phase currents will be equal to one another and shifted through the same angle in phase relative to the respective phase voltages—this is a symmetrical system of phase currents. Now the current in the neutral wire, equal to the vectorial sum of the phase currents, will be zero. Of course, no neutral wire need actually be installed.

Combining the phase current vectors $\bar{I}_A$ and $\bar{I}_B$ (Fig. 7-14) gives a vector equal in magnitude but opposite in direction

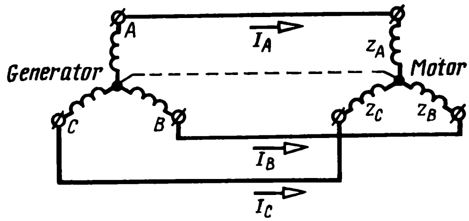


Fig. 7-15. Star connections for a generator and a load

to the vector $\overline{I}_C$. Mathematically, it is written

$$\overline{I}_A + \overline{I}_B = -\overline{I}_C$$

The sum of the three current vectors is then zero:

$$\overline{I}_A + \overline{I}_B + \overline{I}_C = 0$$

Solution of a symmetrical three-phase network reduces to solving a single phase.

If in a star-connected load (Fig. 7-15) the phase impedances are all the same, the phase voltage will be

$$V_p = V_l/\sqrt{3} \tag{7-9}$$

and the phase current,

$$I_p = V_p/z_p \tag{7-10}$$

The cosine of the phase difference between the phase current and voltage is given by

$$\cos \varphi_p = r_p/z_p \tag{7-11}$$

and the sine and tangent of the same phase difference are

$$\sin \varphi_p = x_p/z_p, \quad \tan \varphi_p = x_p/r_p \tag{7-12}$$

The active, reactive and apparent powers per phase are given by

$$P_p = V_pI_p \cos \varphi_p, \quad Q_p = V_pI_p \sin \varphi_p, \quad S_p = V_pI_p \tag{7-13}$$

In a symmetrical system of voltages and currents, the powers in the three phases are given by

$$P = 3P_p = 3V_pI_p \cos \varphi_p \qquad (7\text{-}14)$$

$$Q = 3Q_p = 3V_pI_p \sin \varphi_p \qquad (7\text{-}15)$$

$$S = 3S_p = 3V_pI_p \qquad (7\text{-}16)$$

If we recall that for a star-connected load, $I_p = I_l$ and $V_p = V_l/\sqrt{3}$, then the active power is

$$P = 3V_pI_p \cos \varphi_p = (3/\sqrt{3})\, V_lI_l \cos \varphi_p$$
$$= \sqrt{3}\, V_lI_l \cos \varphi_p \qquad (7\text{-}17)$$

the reactive power is

$$Q = \sqrt{3}\, V_lI_l \sin \varphi_p \qquad (7\text{-}18)$$

and the apparent power is

$$S = \sqrt{3}\, V_lI_l \qquad (7\text{-}19)$$

In an unsymmetrical system of voltages or in an unbalanced system of load phases, the active and reactive powers in a three-phase system are the sums of the respective powers in the individual phases.

Example 7-3. Given: A three-phase, star-connected generator with a phase voltage of 220 V and a load with a resistance of 6 ohms and an inductive reactance of 8 ohms.

To find: The line voltage, the line and phase currents, and the active power in the load.

Solution.

The line voltage is given by

$$V_l = \sqrt{3}\, V_p = 1.73 \times 220 \text{ V} = 380 \text{ V}$$

The impedance per load phase is

$$z_p = \sqrt{r_p^2 + x_p^2} = \sqrt{6^2 + 8^2} = 10 \text{ ohms}$$

The phase current is

$$I_p = V_p/z_p = 220 \div 10 = 22 \text{ A}$$

In a star-connected load, the phase current is equal to the line current, $I_l = 22$ A.

The cosine of the phase difference between the phase current and voltage is

$$\cos \varphi_p = r_p/z_p = 6 \div 10 = 0.6$$

The active power in a three-phase network is

$$P = \sqrt{3}\, V_l I_l \cos \varphi = 1.73 \times 380 \times 22 \times 0.6 = 8.7 \text{ kW}$$

Example 7-4. Given: A star-connected three-phase motor connected to a 380-V supply line and taking 10 kW at $\cos \varphi =$ $= 0.8$. To find: The motor current.

Solution.

The power in the motor circuit is

$$P = \sqrt{3}\, V_l I_l \cos \varphi$$

so the current in the motor circuit is

$$I_l = P/\sqrt{3}\, V_l \cos \varphi = 10{,}000 \div (1.73 \times 380 \times 0.8) = 19 \text{ A}$$

7-5. Delta-Connected Loads

In a delta-connected load (Fig. 7-16), the load phases are connected to the line wires of a generator. As a result, each load phase is directly connected to the line voltage which is at the same time the phase voltage:

$$V_A = V_{AB}, \quad V_B = V_{BC}, \quad V_C = V_{CA}$$

In this case, the phase voltages (in contrast to a star connection) are independent of the phase impedances in the load.

The positive direction for the phase currents is assumed to be from A' to B', from B' to C' and from C' to A'. The positive direction for the line currents is assumed to be from the generator to the load.

By Kirchhoff's first (current) law, the instantaneous currents at point A' are given by

$$i_A + i_{CA} = i_{AB}$$

Hence,

$$i_A = i_{AB} - i_{CA}$$

Similarly, for point B',

$$i_B = i_{BC} - i_{AB}$$

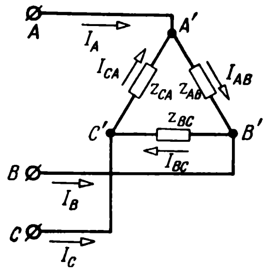


Fig. 7-16. Loads connected into a delta

and for point C',

$$i_C = i_{CA} - i_{BC}$$

To sum up, the *instantaneous line current is equal to the algebraic difference of the instantaneous currents in the phases connected to the given line wire.*

It follows from the above statement that the *line current vector is equal to the difference between the respective phase current vectors* or, mathematically,

$$\bar{I}_A = \bar{I}_{AB} - \bar{I}_{CA}, \quad \bar{I}_B = \bar{I}_{BC} - \bar{I}_{AB}, \quad \bar{I}_C = \bar{I}_{CA} - \bar{I}_{BC} \tag{7-20}$$

In the diagram of Fig. 7-17, the line current vectors are derived as the differences between the respective phase current vectors, with all vectors starting out from a common origin. Sometimes, to make the representation more graphic, the vectors are displaced parallel to themselves so that the voltage vectors form a closed triangle (Fig. 7-18).

In the case of a balanced phase load, namely

$$z_{AB} = z_{BC} = z_{CA} = z_p$$

and

$$\varphi_{AB} = \varphi_{BC} = \varphi_{CA} = \varphi_p$$

the rms phase currents are all the same and displaced through the same phase angle, that is, 120°, relative to their respective voltages (Fig. 7-19). As a result, the phase currents form a symmetrical system. If we drop a perpendicular from the terminal point of an arbitrarily chosen phase current vector, $\bar{I}_{AB}$, onto the line current vector, $\bar{I}_A$, we shall

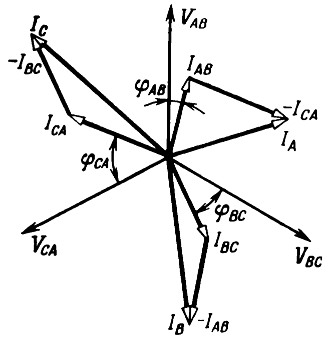


Fig. 7-17. Vector diagram for delta-connected loads

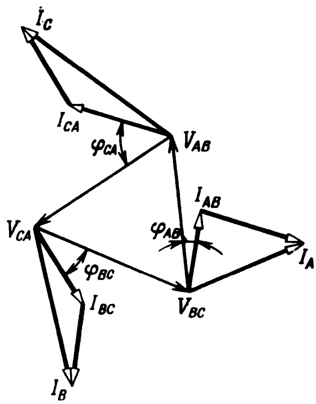


Fig. 7-18. Vector diagram for delta-connected loads

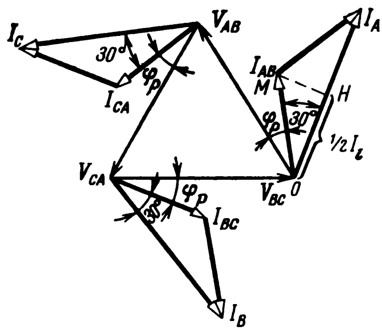


Fig. 7-19. Vector diagram for a delta-connected, balanced-phase network

obtain from the right triangle OMH that

$$I_l/2 = I_p \cos 30° = I_p (\sqrt{3}/2)$$

whence

$$I_l = \sqrt{3}\, I_p \qquad (7\text{-}21)$$

which states that a line current is $\sqrt{3}$ times the respective phase current.

From the same diagram it follows that the line currents lag behind the respective phase currents by 30°.

Solving a symmetrical, delta-connected three-phase system reduces to solving a single phase.

More specifically, the phase voltage is

$$V_p = V_l$$

the phase current is

$$I_p = V_p/z_p$$

and the line current is

$$I_l = \sqrt{3}\, I_p$$

The phase difference between the phase current and voltage in terms of its cosine, sine or tangent is

$$\cos \varphi_p = r_p/z_p, \quad \sin \varphi_p = x_p/z_p, \quad \tan \varphi_p = x_p/r_p$$

The active, reactive and apparent powers of single phase are respectively given by

$$P_p = V_p I_p \cos \varphi_p; \quad Q_p = V_p I_p \sin \varphi_p; \quad S_p = V_p I_p$$

In a symmetrical system of voltages and currents, the respective powers in the three phases may be written

$$P = 3P_p = 3V_p I_p \cos \varphi_p = \sqrt{3}\, V_l I_l \cos \varphi_p \qquad (7\text{-}22)$$

$$Q = \sqrt{3}\, V_l I_l \sin \varphi_p \qquad (7\text{-}23)$$

$$S = \sqrt{3}\, V_l I_l \qquad (7\text{-}24)$$

In the case of an unsymmetrical system of voltages or an unbalanced-phase load, the active and reactive powers of a three-phase system are the sums of the powers in the individual phases and are given by the same expressions as have been derived for star-connected loads.

Delta connection can be used for both lamps (Fig. 7-20) and motors. When using this form of connection, it is essential that the lamps should have a rated voltage equal to the line voltage of the supply mains. A three-phase motor can be delta-connected, if its rated phase voltage is equal to the line voltage of the supply mains, or star-connected, if its rated phase voltage is $1/\sqrt{3}$ of the line voltage of the supply mains.

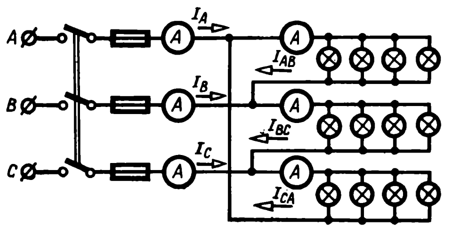


Fig. 7-20. Connection of lamps into a delta

Example 7-5. Given: A three-phase, delta-connected motor operating on 220 V at $\cos\varphi = 0.8$ and dissipating 3 kW. To find: The line and phase currents.

Solution. From Eq. (7-22) it follows that

$$I_l = P/\sqrt{3}\,V_l \cos\varphi = 3000 \div (1.73 \times 220 \times 0.8) = 10 \text{ A}$$

Hence, the phase current is

$$I_p = I_l/\sqrt{3} = 10 \div 1.73 = 6 \text{ A}$$

Example 7-6. Given: A three-phase, delta-connected motor operating on 120 V at 25 A and dissipating 3 kW. To find: The power factor.

Solution.

From Eq. (7-22) it follows that

$$\cos\varphi = P/\sqrt{3}\,V_l I_l = 3000 \div (1.73 \times 120 \times 25) = 0.58$$

Example 7-7. Given: Loads $r_{AB} = 10$ ohms, $r_{BC} = r_{CA} = 20$ ohms connected to a three-phase mains supply (Fig. 7-20) with $V_l = 120$ V. To find: the load voltage in the case of a blown fuse in wire B.

Solution.

Should the fuse blow, the loads AB and BC will be connected in series to a line voltage of $V_l = 120$ V. Hence, the load current will be

$$I_{AB} = I_{BC} = V_{AC}/(r_{AB} + r_{BC}) = 120 \div (10 + 20) = 4 \text{ A}$$

The terminal voltages at the loads will be

$$V'_{AB} = I_{AB} r_{AB} = 4 \times 10 = 40 \text{ V}$$

$$V'_{BC} = I_{BC} r_{BC} = 4 \times 20 = 80 \text{ V}$$
$$V_{CA} = V_l = 120 \text{ V}$$

With any connection of loads, the algebraic sum of the instantaneous line currents in a three-wire three-phase network is zero.

For a star connection with the neutral wire ungrounded, on assuming that the positive direction for the line currents is from generator to load, we can write by Kirchhoff's current law

$$i_A + i_B + i_C = 0 \qquad (7\text{-}25)$$

For a delta connection, the sum of the line currents is zero, because

$$i_A + i_B + i_C = i_{AB} - i_{CA} + i_{BC} - i_{AB} + i_{CA} - i_{BC} = 0$$

Naturally, the sum of the line current vectors is zero

$$\overline{I}_A + \overline{I}_B + \overline{I}_C = 0 \qquad (7\text{-}26)$$

This is the reason why, for example, the mmf due to the three line currents in a three-phase cable is zero, as is the magnetic flux in that cable. This enables one to use steel armour in order to protect the cable against mechanical damage without running any risk of overheating it by reversal of magnetization, as would be the case if the sum of currents were not equal to zero.

Chapter Eight Electrical Measurements and Instruments

8-1. Basic Definitions

Measurement is essentially a comparison of any given quantity with another quantity of the same kind chosen as a standard or a unit, in order to find its value or magnitude. The value or magnitude thus found can be expressed as a dimensional number made up of a numeral and the name of the unit used. For example, a current of 15 amperes.

A *standard* is a measurement facility intended to embody the specified magnitude of a physical quantity.

A *measuring instrument* is a facility intended to generate a signal to represent the result of measurement in a form that the observer can perceive.

Both standards and measuring instruments (or, simply, instruments) may be classed into *reference* and *working*.

Reference standards and instruments are resorted to in order to check or verify working measuring facilities. As a result of such a check or verification, the magnitude of the quantity (or quantities) involved is transferred to the working standard or instrument in question with a definite degree of accuracy. Reference standards and instruments are established as such by a nation's metrological authority.

Working standards, as their name implies, are used for practical measurements, that is, where the transfer of the size of a unit to other standards or instruments is not involved.

In any measurement, the result somewhat differs from the true value of the quantity being measured. By custom, the true value of a quantity is that which is found by means of a reference standard and instrument. *The difference between*

the true value and that actually measured is called the absolute error of measurement.

The quality of measurement is ordinarily stated in terms of the percent-of-indication accuracy or error. It is defined as the percent ratio of the absolute error to the true or actually measured value of the quantity.

Example 8-1. Given: The value of current actually measured is $I_{ind} = 26$ A; its true value is $I = 25$ A. To find: The absolute and percent-of-indication errors.

Solution.

The absolute error in the current measured is

$$\Delta I = I_{ind} - I = 26 - 25 = 1 \text{ A}$$

The percent-of-indication error is

$$\gamma_I = (\Delta I / I) \times 100\% = (1/25) \times 100\% = 4.0\%$$
$$= \text{(approx.)}\ (1/26) \times 100\% = 3.8\%$$

8-2. Classification of Electrical Measuring Instruments

Electrical measuring instruments may be classed into two broad groups as follows:

(1) Direct-reading instruments which display the numerical value of the quantity being measured on their dial. Examples are an ammeter and a voltmeter.

(2) Comparison instruments which find the value of the unknown through a comparison with a standard. An example is a bridge.

According to the manner in which they display the value measured, instruments are divided into analog instruments (such as pointer meters) and digital instruments which present a digital readout.

According to the electrical quantity being measured (current, voltage, power, frequency, phase difference, resistance, energy and the like), instruments are classed into ammeters, voltmeters, wattmeters, frequency meters, phase meters, ohmmeters (or megohmmeters), energy meters, etc.

For their operation, instruments may depend on various principles, as will be clear from reference to Table 8-1. The choice of a particular type of instrument depends on

Table 8-1

Some of the Instruments and Their Diagram Symbols

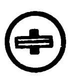
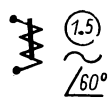
Type symbol	Type name	Markings on dials	Particulars
	Moving-coil	Accuracy classes	
		0.05 0.1 0.2 0.5 1 1.5 2.5 4	Basic percent-of-full-scale accuracy: 0.05, 0.1, 0.2, 0.5, 1.0, 1.5, 2.5 and 4%
	Moving-coil ratiometer	Current	
	Rectifier-type	—	d. c.
		~	a. c.
	Thermo-emf	≋	Three-phase
		Mounting position	
	Moving-iron	⊥	Vertical dial position
		⊓	Horizontal dial position
	Electrodynamic	∠	Inclined dial position
		Insulation strength	
	Electrodynamic ratiometer		Measuring circuit insulated from case and tested for a voltage of 2 kV
		Terminals	
	Ferrodynamic	∗	Generator terminal
		⊥	Case terminal
	Induction	⏚	Grounding terminal

Table 8-1 (cont.)

Type symbol	Type name	Markings on dials	Particulars
	Electrostatic	Example 	Moving-iron instrument, accuracy class 1.5, for a.c., with a dial inclined at 60° from the horizontal
	Vibration		

how closely its performance (or properties) answer the requirements of measurements and fit the conditions in which they are used.

According to USSR State Standard GOST 1845-59, electrical measuring instruments are divided into eight accuracy classes, and each class is assigned an accuracy rating in the form of a dimensionless number, namely 0.05, 0.1, 0.2, 0.5, 1.0, 1.5, 2.5, and 4. On the dials of the instruments, the accuracy rating is shown by the respective number in a circle.

The accuracy rating defines *the limit, expressed as a percentage of full-scale value, that basic errors will not exceed when the instrument is used under reference conditions. The numerator of the ratio is the absolute error,* Δx, *and the denominator of the ratio or, as it is technically called, the base value, is the full-scale deflection (f.s.d.) of the instrument,* x_{fsd}.

Hence, the percent-of-full-scale error is

$$\gamma_{fsd} = (\Delta x / x_{fsd}) \times 100\% \qquad (8\text{-}1)$$

The term 'reference conditions' applies to the use of the instrument in a physical position stated on the dial, normal ambient temperature (usually, +20°C), and the absence of any stray electromagnetic fields (except the terrestrial field).

The percent-of-indication error in the value of the unknown, x_{ind} may be defined as the ratio of the maximum absolute error Δx of the instrument to the true (or, in approximate

terms, to the actually measured) value of the quantity, x_{ind}

$$\gamma_{ind} = (\Delta x / x_{ind}) \times 100\% \tag{8-2}$$

On replacing Δx in Eq. (8-2) by $\Delta x = \gamma_{fsd} x_{fsd}/100\%$ from Eq. (8-1), we obtain

$$\gamma_{ind} = (\Delta x / x_{ind}) \times 100\% = (\gamma_{fsd} x_{fsd}/100\% \, x_{ind}) \times 100\% = \gamma_{fsd} x_{fsd}/x_{ind} \tag{8-3}$$

Thus, *the percent-of-indication error of measurement is equal to the instrument error multiplied by the ratio of the full-scale value (instrument range) to the true or measured value of the unknown quantity.*

The error in the value of a quantity as measured by a given instrument decreases as the value measured approaches the full-scale value (range) of the instrument. Hence, in order to use an instrument better, it should be employed to measure values likely to fall in the second half of the scale.

The errors of measurement and instruments may be both positive and negative.

Example 8-2. Given: A 1.5-accuracy class ammeter with a range (f.s.d. value) of 30 A and an indication of 15 A.

To find: The percent-of-indication accuracy (or error) of the instrument.

Solution.

The percent-of-indication error in current measurement for this instrument is

$$\gamma_I = \gamma_{fsd} (I_{fsd}/I_{ind}) = \pm 1.5\% \, (30/15) = \pm 3\%$$

Electrical measuring instruments are expected to comply with a multitude of requirements, the chief among which are:

1. The errors of the instrument ought not to exceed the limits set by the relevant standard for the accuracy class to which the instrument belongs.
2. The power lost in the instrument should be as small as practicable.
3. The instrument scale must preferably be uniform.

4. The oscillations arising in the instrument must be well damped and the instrument should have a reliable insulation.
5. The instrument should stand up to overloads well.

8-3. The Movement of an Instrument

The key part of any direct-indicating instrument is its movement.

More specifically, the movement incorporates those parts that move as a direct result of a variation in the quantity that the instrument is measuring. The motion of these parts is usually converted to the angular deflection of a pointer which indicates the value of the unknown.

(a) The Moving-Coil Movement

Moving-coil movements may be of two types, one with an external and the other with an internal permanent magnet (Fig. 8-1).

Referring to Fig. 8-1, this type of movement contains a stationary steel cylinder B, a permanent magnet NS with pole-pieces N' and S', an air gap A, and a rectangular moving coil C, supported by two pivots which are in turn carried in two bearings. The front pivot carries a pointer whose tip moves over a scale drawn on a dial (or scale-plate, as it is sometimes called). The coil consists of a former and a winding wound with fine insulated wire designed for a nominal current of 10 to 100 mA.

The permanent magnet produces a uniform radial magnetic field in the air gap, which interacts with the current that is fed to the coil by helical springs. This interaction gives rise to a force couple FF (Fig. 8-2) which in turn produces a torque. Under the action of this torque, the coil (the moving element) turns through an angle α, to a position where the torque is balanced by the restoring, or control, torque produced by the springs.

Because the torque is proportional to the current through the coil

$$T = kI$$

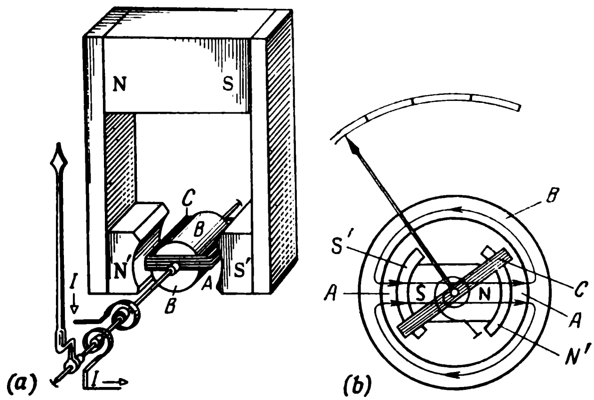


Fig. 8-1. Moving-coil movements with an external permanent magnet (*a*) and with an internal permanent magnet (*b*)

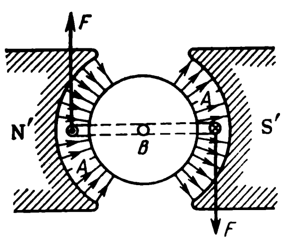


Fig. 8-2. Generation of torque by a moving-coil movement

and the control torque is proportional to the twist angle of the springs

$$T_c = D\alpha$$

we may write

$$T = T_c = kI = D\alpha \tag{8-4}$$

where k and D are proportionality factors.

It follows from Eq. (8-4) that the angle of deflection of the moving element is

$$\alpha = kI/D = S_I I$$

and the coil current is

$$I = D\alpha/k = c_I\alpha \tag{8-5}$$

where $S_I = \alpha/I$ is the current sensitivity of the instrument, expressed by the number of scale divisions through which the pointer deflects in response to a change of one unit in the coil current, and $c_I = D/k = I/\alpha$ is the current constant known for each instrument.

Thus, the value of the unknown current can be defined as the product of the deflection angle (read on the scale) and the current constant c_I.

Each variation in the unknown quantity causes the moving element to oscillate. It is customary to minimize, or dampen, such oscillations. The damping action is provided by a *damper* (often called a dash-pot).

Moving-coil movements utilize magnetic-induction damping. The damper is the coil former. As the coil rotates, the magnetic flux linked by the former is varied. The interaction of the current induced in the former with the magnetic field produces a *retarding torque* which provides the damping action.

In the case of an alternating current, the torque is proportional to the instantaneous value of current. At the standard frequency, the torque oscillates so rapidly that the moving element has time only to turn through an angle proportional to the torque averaged over a cycle and, as a consequence, to the current averaged over a cycle. However, a sinusoidal current averaged over a cycle is zero, so the moving element does not deflect. Hence, the moving-coil movement examined above is only suitable to measure direct current.

If, however, the coil placed between the poles N and S of a permanent magnet, *1* (Fig. 8-3) is made in the form of a narrow loop, *2*, the latter will have an insignificant inertia, and the alternating current flowing through the loop will cause its middle part carrying a mirror, *3*, to turn through an angle proportional to the instantaneous value of current. The amplitude of deflection of the light beam reflected from the mirror will then give a measure of the current in the loop.

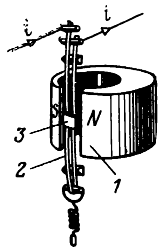


Fig. 8-3. Vibrator of a mirror galvanometer oscillograph

This form of instrument, ordinarily referred to as a mirror galvanometer, is incorporated in mirror-galvanometer oscillographs to record or display (for visual observation) current variations.

(b) A Moving-Iron Movement

A moving-iron movement (Fig. 8-4) has a fixed coil, A, a movable iron core, B, mounted on a shaft together with a pointer, a spring and an aluminium sector damper, C. The unknown current flowing in the coil magnetizes the core, and the latter is pulled in. This action causes the pointer to deflect and indicate the strength of the current.

When the moving element rotates, eddy currents are induced in the sector damper C located in the magnetic field set up by a magnet M. The eddy currents interact with the magnetic field of the same magnet and produce a retarding torque which supplies the damping action.

Moving-iron instruments can measure both direct and alternating currents, because the core will be pulled inside the coil with the current flowing in any direction.

Because of the residual magnetization of the core, the angle of deflection may be different for the same value of current on the rising and falling portions of the current waveform. This results in a complementary residual-magnetization error. To minimize it, the core is fabricated from a material having a low residual induction (such as Permalloy).

The effect of external stray magnetic fields on a moving-iron movement can be minimized in one of several ways,

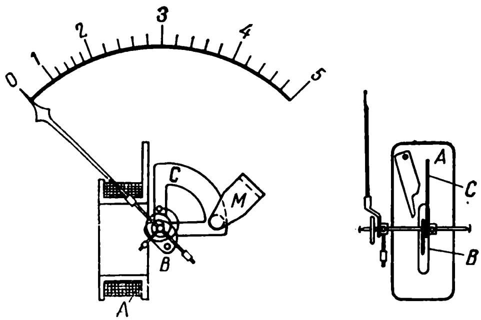


Fig. 8-4. Moving-iron meter movement

namely: (1) by enclosing the movement in a steel shield or enclosure; (2) by using an astatic movement with two cores on a common shaft and two coils connected in series. The coils energized by the same unknown current set up fields acting in opposite directions. In the same manner, an external stray uniform field will buck the field of one coil and boost that of the other, with the result that the net effect is sufficiently small.

(c) An Electrodynamic Movement

An elementary electrodynamic movement (Fig. 8-5) has two coils, one fixed, A, and the other movable, B. The movable coil, pointer, the vane C of an air dash-pot and the ends of two springs are attached to a common shaft. The electrodynamic interaction between the currents I_1 and I_2 flowing in the coils gives rise to a torque (Fig. 8-6). This torque causes the moving element to turn through an angle α to a position in which it is balanced by the control or restoring torque supplied by the springs.

In a d.c. circuit, the torque and the deflection angle of the moving element are proportional to the product of the two currents

$$\alpha = k_1 I_1 I_2 \tag{8-6}$$

In an a.c. circuit, the instantaneous torque is proportional to the product of the instantaneous currents, and the torque

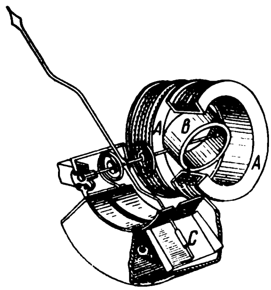


Fig. 8-5. Electrodynamic meter movement

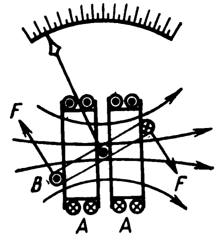


Fig. 8-6. Generation of torque by an electrodynamic meter movement

and deflection angle averaged over a cycle are dependent not only on the instantaneous currents, but also on the cosine of the phase difference between them

$$\alpha = k_1 I_1 I_2 \cos \psi \tag{8-7}$$

The deflection angle gives a measure of the unknown quantity.

The weak magnetic field established in the moving element produces a correspondingly weak torque, so, in order to retain the high accuracy inherent in this type of movement, it is important to minimize the errors arising from friction in the bearings. This is achieved by reducing the mass of the moving element and by giving a high surface finish to the pivots and bearings.

The effect of stray external magnetic fields is minimized by shielding or by the use of an astatic movement. Electrodynamic movements are sensitive to overloading.

(d) A Ferrodynamic Movement

A ferrodynamic movement (Fig. 8-7) has a magnetic circuit, A, and a fixed cylindrical core, B, fabricated from electrical-sheet steel laminations. The current in a coil, C, establishes a magnetic flux which interacts with the current in a moving coil, D, mounted on the same shaft, or pivot, with a pointer. Thus, a ferrodynamic instrument operates by the same principle as an electrodynamic instrument. Accordingly, Eqs. (8-6) and (8-7) remain in force.

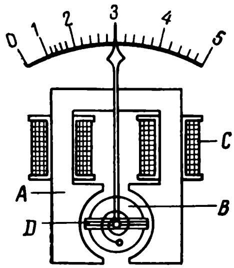


Fig. 8-7. Ferrodynamic meter movement

The presence of steel augments the magnetic flux and the torque, so a more robust construction is achieved. Stray extraneous magnetic fields have practically no effect on indications of ferrodynamic instruments.

8-4. Measurement of Current and Voltage

Indications of an ammeter depend on the current I_A passing through the instrument, so in order to measure the current in any load, I_L, the ammeter is connected in series with the load so that $I_A = I_L$ (Fig. 8-8).

The presence of an ammeter in the circuit ought not to affect the current being measured, so its resistance must be small in comparison with that of the load. The small resistance of the ammeter, r_A, entails a low nominal power dissipation in it:

$$P_{An} = I_{An}^2 r_A$$

If the direct current to be measured exceeds the range (full-scale value) of an ammeter, one uses a shunt (see Sec. 8-4*a*). In the case of an alternating current, resort is made to instrument current transformers (see Sec. 9-11).

A voltmeter measures the voltage applied across its terminals, V_V, so in order to measure the voltage across any load, V_L, the terminals of the voltmeter must be connected to those of the load as shown in Fig. 8-9. Then $V_V = V_L$.

Indications of a voltmeter depend on its current I_V. If they are to depend uniquely on the voltage V_V, the voltmeter resistance must be constant, because it is only then

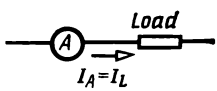


Fig. 8-8. Connection of an ammeter in circuit

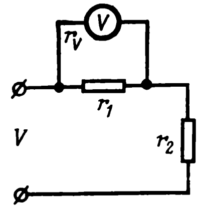


Fig. 8-9. Connection of a voltmeter in circuit

that $V_V = I_V r_V = I_V r_{const}$. Similarly, for the nominal values,

$$V_{Vn} = I_{Vn} r_V$$

that is, the voltmeter resistance is proportional to the voltmeter range (full-scale value).

The presence of a voltmeter in the circuit ought not to affect the unknown voltage, so the resistance of the voltmeter must be high in comparison with the load resistance with which it is connected in parallel. If the voltmeter resistance r_V is high, the nominal current of the voltmeter will be low, and so will its nominal power loss, because

$$I_{Vn} = V_{Vn}/r_V$$

and

$$P_{Vn} = V_{Vn} I_{Vn} = V_{Vn}^2/r_V$$

The voltmeter resistance can be held at a practically constant value and independent of temperature variations by placing a high-value, manganin resistor, r_s, in series with the voltmeter (hence, the name 'series resistor'). Since the value of r_s remains practically unchanged, we may write

$$r_m + r_s = r_V = \text{constant}$$

(a) Moving-Coil Ammeters and Voltmeters

Moving-coil instruments for small currents — galvanometers, microammeters and milliammeters — have their coil connected to the terminals and their scale calibrated in the respective units of current.

A moving-coil ammeter (see Fig. 8-1) has its movement connected in parallel with a resistor, called a shunt (Fig. 8-10),

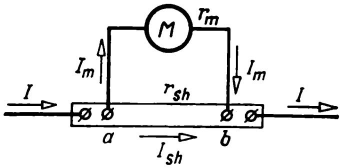


Fig. 8-10. Meter movement and a shunt

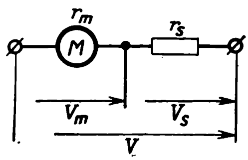


Fig. 8-11. Meter movement and a series resistor

whose purpose is to extend the range of the instrument, that is, to raise its nominal current.

The unknown current I is divided into the shunt current I_{sh} and the movement current I_m. The voltage across this parallel circuit V_{ab} (see Fig. 8-10) is given by

$$I_m r_m = I \frac{r_m r_{sh}}{r_m + r_{sh}}$$

Hence, the unknown current is

$$I = I_m \frac{r_m + r_{sh}}{r_{sh}} = I_m p \tag{8-8}$$

If the movement and shunt resistances, r_m and r_{sh}, are held constant, there will be an unvarying relation between the currents I and I_m, so the deflection angle of the pointer can give a measure of the unknown current I.

Use is made of built-in shunts which are enclosed in the instrument, and external shunts placed outside the instrument.

A moving-coil voltmeter has its movement (see Fig. 8-1) connected to a series resistor (Fig. 8-11) whose purpose is to hold the voltmeter resistance constant and, also, to extend its range*. The scale of a voltmeter is calibrated in units of voltage

$$V_V = I_m (r_m + r_s)$$

The ratio of the voltage across the terminals of a voltmeter, V_V, to that across the terminals of its movement, $V_m =$

* Because of its second function, the series resistor of a voltmeter is often called a multiplier resistor.— *Translator's note.*

$= I_m r_m$, is constant

$$p = V_V/V_m = (r_m + r_s)/r_m \tag{8-9}$$

Working voltmeters use single-range series resistors; reference and laboratory voltmeters use multirange series resistors so that any one of several ranges can be selected by bringing in circuit a particular part of the resistor.

In a.c. circuits, in addition to series resistors, the range of a voltmeter can be extended by using an instrument voltage transformer (see Sec. 9-11).

Moving-coil instruments are fabricated in accuracy classes 0.1 through 2.5.

Of the properties of these instruments the following may be noted. They can be used in d.c. circuits, have high sensitivity, are insensitive to temperature variations and external magnetic fields, have a uniform scale, dissipate little power, but are sensitive to overloading.

(b) Rectifier-Type Ammeters and Voltmeters

A rectifier-type ammeter consists of a moving-coil movement (see Fig. 8-1) and a semiconductor rectifier (which is an a.c.-to-d.c. converter). A rectifier-type voltmeter additionally has a series resistor.

In an elementary case, a rectifier-type ammeter (Fig. 8-12) consists of a movement connected in series with a crystal diode (see Sec. 16-3) so that only one half-cycle of alternating current passes through its movement during each cycle. The other half-cycle of current passes through a second diode connected in reverse direction. The torque averaged over a cycle and the deflection of the moving element depend on the average current through the movement, which in the case of a sinusoidal waveform is proportional to the rms value of current. It is the rms values that are marked on the scale of an ammeter. The range can be extended through the use of shunts.

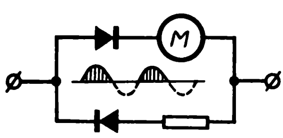


Fig. 8-12. Circuit of a simple rectifier-type ammeter

In a voltmeter, because its resistance is held constant, the rms current is proportional to the rms value of its terminal voltage, and it is the latter voltage that is indicated on the scale of the instrument.

Rectifier-type instruments are intended for use in a.c. circuits at frequencies up to 10 kHz. Their accuracy class is 1.5 to 2.5.

(c) Thermo-EMF Ammeters and Voltmeters

A thermo-emf ammeter consists of a moving-coil movement and a thermocouple transducer which may be either of the contact type (Fig. 8-13*a*) or of the insulated type (Fig. 8-13*b*). A thermo-emf voltmeter additionally has a series resistor.

The subdivision into the contact type and the insulated type refers to the manner in which the thermocouple, T, is connected to its heater, H (see Fig. 8-13). In the former type, the ends of the couple are welded directly to the heater. In the latter type, the thermocouple is held close to, but not touching, the heater. The thermocouple is formed by two wires of dissimilar metals, the hot ends of which are welded together and the cold ends are connected to the instrument movement.

On traversing the heater, the unknown current raises its temperature, and the hot ends of the thermocouple are heated. As a result, a thermo-emf appears at the cold ends of the thermocouple, and a current begins to flow in the coil of the movement, thereby causing the coil to turn through an angle proportional to the unknown current. The scale of an ammeter is calibrated in units of current, and that of a voltmeter

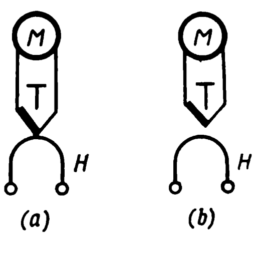


Fig. 8-13. Thermo-emf ammeters

in units of rms voltage which is proportional to the rms current owing to the constancy of the voltmeter resistance.

Thermo-emf instruments are intended for use in a.c. circuits at frequencies up to 50 MHz. They come in accuracy classes 1.5 through 2.5.

(d) Moving-Iron Ammeters and Voltmeters

A moving-iron ammeter is a moving-iron movement (see Fig. 8-4) with a scale calibrated in values of current flowing in its coil. The coil of a moving-coil ammeter is wound with a wire whose cross-sectional area corresponds to the nominal current.

A moving-iron voltmeter consists of a movement (see Fig. 8-4) for a nominal current of 20 to 30 mA, a series manganin resistor, and a scale calibrated in units of voltage.

Because the resistance of a series resistor is many times the reactive impedance of the coil of the movement, the voltmeter has a practically resistive impedance, independent of temperature and frequency.

The deflection of the moving element depends on the current in the coil and the terminal voltage proportional to that current.

Moving-iron instruments are intended for use in a.c. circuit at power (commercial) frequency. They are available in accuracy classes 0.5 through 2.5. They are widely used for engineering measurements owing to their low cost, simplicity and reliability.

(e) Electrodynamic and Ferrodynamic Ammeters and Voltmeters

An electrodynamic ammeter consists of an electrodynamic movement (see Fig. 8-5) in which the coils are connected in parallel and the scale is calibrated in the values of current passing through the ammeter.

To reduce the friction error, the movable coil of the movement is made light in weight and wound with fine wire. The fixed coil is wound with a wire whose cross-section depends on the nominal current of the ammeter. In milliammeters, the coils are connected in series (see Fig. 8-14*a*).

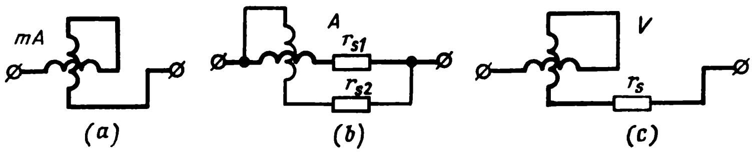


Fig. 8-14. Connections of the coils in an electrodynamic milliammeter, ammeter and voltmeter

When the coils are connected in series, they carry the unknown current $I_1 = I_2 = I$, and the deflection of the moving element is proportional to the square of the current

$$\alpha = k_1 I_1 I_2 \cos \Psi = k_2 I_2 \tag{8-10}$$

When the coils of the ammeter are connected in parallel, the deflection angle is proportional to the square of the current, as defined by Eq. (8-10), if the resistances of the series resistors, r_{s1} and r_{s2}, in the branches are chosen such that the branch currents I_1 and I_2 are in phase with each other ($\psi = 0$), and each is proportional to the unknown current, I.

An electrodynamic voltmeter consists of an electrodynamic movement (see Fig. 8-5) in which the coils (for a nominal current of 20 to 50 mA) are connected to a series resistor (Fig. 8-14*c*) whose functions are to extend the range of the instrument and to minimize the effect of temperature variations, type of current and frequency on voltmeter readings.

Electrodynamic instruments are manufactured in accuracy classes from 0.1 to 0.5 for use in d.c. and a.c. circuits at power and elevated frequencies (up to 2 kHz). They are sensitive to overloading and stray magnetic fields. To minimize the effect of stray magnetic fields, resort is made to shields and astatic movements.

Ferrodynamic ammeters and voltmeters are mainly used in a.c. circuits as component parts of recorders. Their circuits are arranged in the same manner as those of electrodynamic instruments. They produce an increased torque and are robust and reliable in construction, and insensitive to stray magnetic fields. They come in accuracy classes 1.5 to 2.5.

(f) Digital Instruments

A digital instrument is a measuring device which converts a continuous (analog) quantity to a digital quantity and displays it on a readout device as a number with three or four significant digits.

Digital instruments may be classed into two general groups, electromechanical in which the analog-to-digital conversion is performed by some electromechanical device, and electronic where this conversion is done by pulse circuits.

Of the instruments in the first group, use is most often made of voltmeters and ohmmeter. The most commonly

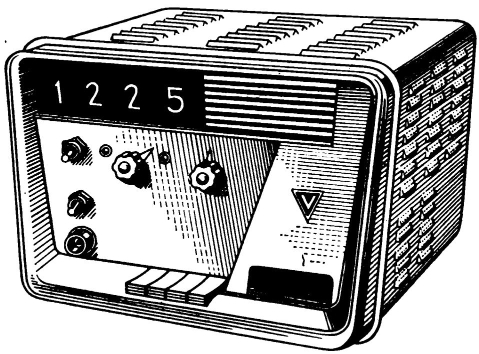


Fig. 8-15. External appearance of the Soviet-made type Щ1411 digital voltmeter

used instruments in the second group are voltmeters, frequency meters and phase meters.

Most often, a digital voltmeter measures the unknown voltage by comparing it with a standard, that is by the null-balance (or, simply, null) method.

When it is applied to the input terminals of a digital instrument, the unknown analog quantity is automatically digitized by a suitable electronic circuit. After the conversion, lamps are turned on in the digital readout unit (see Fig. 15-13) to display the numerical value of the unknown quantity (Fig. 8-15).

Digital voltmeters are manufactured for ranges from 100 μV to 1 kV. The time per measurement does not exceed

1 s for electromechanical instruments and a few milliseconds for electronic ones.

When digital instruments having a high accuracy (their error is as small as 0.01 to 0.1%) are used in conjunction with printers, one can readily automatize the process of measurement and data presentation. In combination with computers, digital instruments are used for automatic process monitoring and control.

Among the disadvantages of digital instruments are complex design and high cost.

8-5. Power Measurements

Once we have measured the voltage across and the current in a d.c. circuit, V and I, respectively, we can readily find the power in that circuit by the equation

$$P = VI \tag{8-11}$$

This power can alternatively be measured by an electrodynamic wattmeter.

An electrodynamic wattmeter consists of an electrodynamic movement and a scale calibrated in units of power. The fixed coil in a wattmeter is called a *current* or *series* coil, because it is connected in series with the load (Figs. 8-16 and 8-17)*.

The movable coil of a wattmeter and a nonreactive manganin series resistor, r_s, make up the *voltage* or *shunt* circuit of the instrument, because it is connected in parallel with the load (Fig. 8-17) in which the power is being measured.

The deflection of the moving element in an electrodynamic wattmeter is proportional to the product of the currents in its coils, given by Eq. (8-6):

$$\alpha = k_1 I I_V$$

Since the resistance of the shunt circuit is constant, its current varies in proportion to the voltage, $I_V = V/r_V$, and the deflection of moving element is proportional to the power

$$\alpha = k_1 I I_V = k_1 (1/r_V) IV = k_2 IV = k_2 P \tag{8-12}$$

* In Fig. 8-17 and elsewhere, the symbol for a wattmeter has two circles. The inner circle designates the movement connecting the two coils, and the outer circle represents the case.

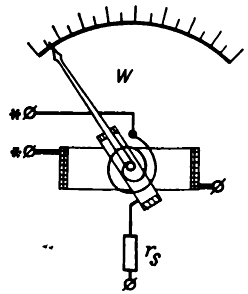


Fig. 8-16. Movement of an electrodynamic wattmeter

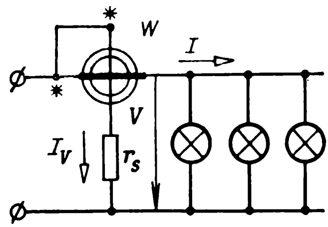


Fig. 8-17. Connection of the movement in an electrodynamic wattmeter

The active power in a.c. circuits, $P = VI \cos \varphi$, is measured by electrodynamic and ferrodynamic wattmeters. In this case, the deflection of the moving element, given by Eq. (8-7), is

$$\alpha = k_1 I I_V \cos \psi$$

Because the current in the shunt circuit, $I_V = V/r_V$, varies in proportion to, and is in phase with, the unknown voltage (Fig. 8-18), the phase difference between the currents in the coils of the instrument is equal to the phase difference φ between the current I and the voltage V, so the deflection of the moving element in a wattmeter is proportional to the active power in the circuit

$$\alpha = k_1 I I_V \cos \psi = k_1 (1/r_V) IV \cos \varphi = k_2 IV \cos \varphi = k_2 P \quad (8\text{-}13)$$

The series-coil terminal connected to the supply source is called a generator terminal; the shunt-circuit terminal connected to the series coil is also called a generator terminal. The generator terminals are marked on the wattmeter by asterisks.

In setting up a circuit, the operator must never interchange the series-coil and shunt-circuit terminals, as this would reverse the direction of current flow or phase reversal of the respective current; as a result, the moving element of the wattmeter would be forced to rotate in the reverse direction.

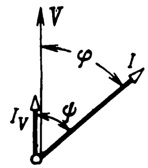


Fig. 8-18. Vector diagram for an electrodynamic wattmeter

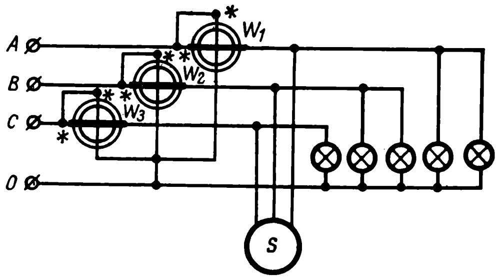


Fig. 8-19. Connection of the movements in three wattmeters to measure power in a four-wire, three-phase network

If the voltage in an a.c. circuit is over 220 V and the current exceeds 5 A, a wattmeter will usually be connected to such a circuit via instrument transformers (see Sec. 9-11).

The active power in a four-wire three-phase network is

$$P = P_A + P_B + P_C = I_A V_A \cos \varphi_A + I_B V_B \cos \varphi_B + I_C V_C \cos \varphi_C \quad (8\text{-}14)$$

It is measured by three wattmeters connected as shown in Fig. 8-19; in this arrangement each wattmeter measures the power in one phase. It is advantageous to use a three-unit wattmeter, that is, one made up of three fixed coils and three movable coils driving a common shaft carrying a pointer. The power in a three-phase network can then be read directly from the dial of the instrument.

The power of a symmetrical three-phase network can be found by measuring the power in one phase by a wattmeter (see Figs. 8-20 and 8-21), $P_w = P_p$, and multiplying it by three:

$$P = 3P_w = 3P_p$$

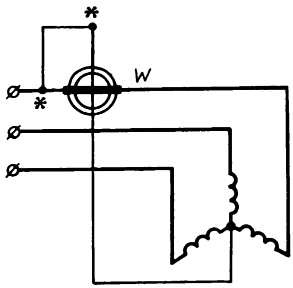


Fig. 8-20. Connection of the movement in wattmeter to measure the power of a motor with the neutral point available for connection

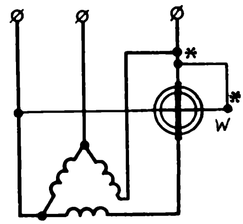


Fig. 8-21. Connection of the movement in a wattmeter to measure the power of a delta-connected motor

The power in a three-wire, three-phase network in the case of balanced and unbalanced load is measured by a two-unit wattmeter.

A two-unit electrodynamic or ferrodynamic wattmeter has two fixed coils and two movable coils mounted on a common shaft carrying a pointer (Fig. 8-22).

The instantaneous power in a three-phase network is the sum of the instantaneous powers in the three phases

$$p = p_A + p_B + p_C = i_A v_A + i_B v_B + i_C v_C \qquad (8\text{-}15)$$

On replacing the current i_C by its expression from Eq. (7-25),

$$i_C = -i_A - i_B \qquad (8\text{-}16)$$

we obtain

$$\begin{aligned} p &= i_A v_A + i_B v_B - i_A v_C - i_B v_C \\ &= i_A (v_A - v_C) + v_B (v_B - v_C) \\ &= v_A v_{AC} + i_B v_{BC} = p_1 + p_2 \end{aligned} \qquad (8\text{-}17)$$

It follows from Eq. (8-17) that the instantaneous power in a network is the sum of two components, p_1 and p_2.

Let a two-unit wattmeter be connected in circuit (Fig. 8-23) in accordance with Eq. (8-17). The series coil of the first unit is interposed in wire A (it will carry the current i_A), and the series coil of the second unit is interposed in wire B

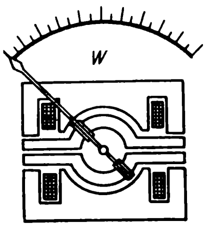


Fig. 8-22. Sketch of a two-unit wattmeter

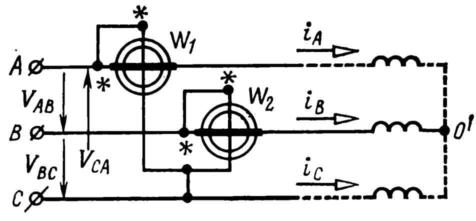


Fig. 8-23. Connection of the movement in a two-unit wattmeter to measure the power in a three-wire, three-phase network

(it will carry the current i_B). The shunt circuit of the first unit is connected to the wires A and C (to carry the voltage v_{AC}), and that of the second to the wires B and C (to carry the voltage v_{BC}). With this arrangement, the instantaneous torque acting on the moving element is proportional to the instantaneous power in the circuit, and the deflection of the moving element, proportional to the average torque, will also be proportional to the average or active power in the three-phase network:

$$P = I_A V_{AC} \cos \varphi_{A\text{-}AC} + I_B V_{BC} \cos \varphi_{B\text{-}BC} \qquad (8\text{-}18)$$

where φ_{A-AC} and φ_{B-BC} are the phase difference between I_A and V_{AC}, and between I_B and V_{BC}, respectively.

The above arrangement holds for any load connection, be it a star or a delta, so any delta connection may be replaced by an equivalent star connection.

Instead of the wires A and B, the series coils of a wattmeter may be connected to any two wires of a three-phase network. The generator terminal of each shunt circuit in

the wattmeter must be connected to the line wire containing the series coil of the respective wattmeter unit. The non-generator terminals of the shunt circuits must be connected to the line wire not containing the series coils of the wattmeter.

A two-unit wattmeter may be replaced by two single-phase wattmeters connected in the same manner (Fig. 8-23). The active power in a three-phase network will then be given by the algebraic sum of their indications. If one of the wattmeters deflects in the reverse direction, the wires connected to the terminals of the shunt circuit in that wattmeter must be interchanged and its indication must be taken with a minus sign.

8-6. Energy Measurement

The electric energy dissipated in single- and three-phase networks is measured by induction meters. In d.c. circuits, use is made of electrodynamic, ferrodynamic and other meters.

An electric meter is a totalizing instrument. Its major distinction from a pointer instrument is that the deflection of its moving element not restrained by a spring is cumulative, and meter indications are added together, or totalized. Each revolution of the moving element corresponds to a precise quantity of energy expended.

An induction meter (Fig. 8-24) consists of an aluminium disc mounted on a shaft, two electromagnets (a series magnet, A, and a shunt magnet, B), a control or restoring magnet, M, and a counter driven by a gear, C.

When the electromagnets are energized, they establish two magnetic fluxes which link the disc and induce in it eddy currents, I_A and I_B. The interaction of I_A with the magnetic flux Φ_B, and of I_B with Φ_A gives rise to a torque T proportional to the power in the load

$$T = k_1 P \tag{8-19}$$

This torque drives the disc of the meter.

As the disc rotates in the field of the control or restoring magnet, eddy currents are induced in the disc (Fig. 8-24). The interaction of these currents with the field of the same

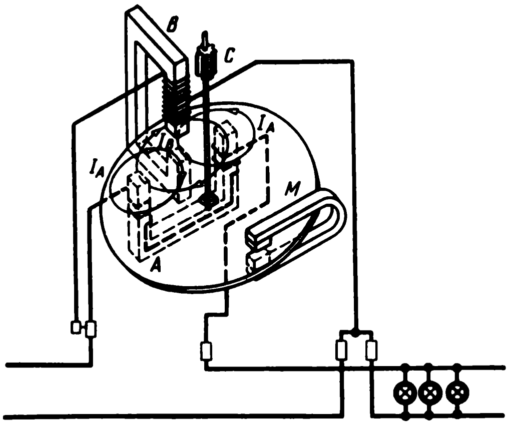


Fig. 8-24. Design and connection of an induction meter

magnet gives rise to a control torque proportional to the rotational speed of the disc, n, namely

$$T_c = k_2 n \tag{8-20}$$

So long as the disc is rotating at a constant speed, its drive and control torques are equal

$$T = T_c$$

or

$$k_1 P = k_2 n$$

Hence, the power in the load is

$$P = (k_2/k_1)\, n = kn$$

Thus, the rotational speed of the disc is proportional to power.

If, in a time t, a network has expended an amount of energy $W = Pt$, the disc will have completed N revolutions in that time, so

$$W = Pt = knt = kN \tag{8-21}$$

Hence, the number of revolutions registered by the counter is proportional to the energy expended.

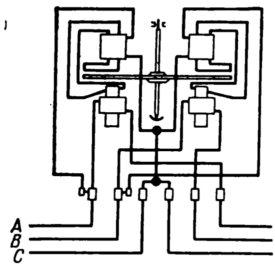


Fig. 8-25. Design and connection of a two-unit, single-disc meter

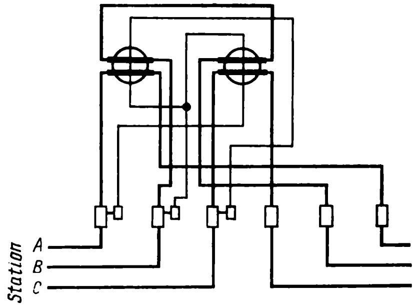


Fig. 8-26. Connection of the SR4-ITR var-hour meter in circuit

The quantity of energy expended by a network during one revolution of the disc

$$W/N = k \tag{8-22}$$

is calleb the constant of the meter.

The energy expended in a network is registered by the counter. To determine the quantity of energy expended over a particular time interval, the indication of the meter at the end of that interval must be subtracted from the indication at the beginning of the same interval.

Active-energy meters in the Soviet Union come in accuracy classes 1, 2 and 2.5, and reactive-energy meters in accuracy classes 2 and 3.

In three-phase, four-wire networks, energy is measured by three-unit meters. It has three electromagnetic systems similar to that of a single-phase meter, each driving a separate disc mounted on a common shaft with a gear that actuates the counter. Such a meter is connected in circuit similarly to a three-unit wattmeter (see Fig. 8-19).

In three-wire three-phase networks, energy is measured by a two-unit two-disc or single-disc meters (Fig. 8-25) or a pair of single-phase meters.

Reactive energy in three-phase networks is measured by var-hour meters, such as the Soviet-made type SR4-ITR (Fig. 8-26).

The SR4-ITR is a two-unit induction meter with two coils on each of the series electromagnets. In their cores, the coils set up magnetic fluxes which, together with the fluxes of the shunt electromagnets, produce torques proportional to the reactive power. The counter registers the reactive energy directly.

The range of wattmeters and meters can be extended by means of instrument transformers (see Sec. 9-11).

8-7. Resistance Measurement

(a) A Resistance Bridge

The standard ordinarily used to measure resistances is a precision resistor. An assembly of precision resistors connected in a particular manner and enclosed in a common case is called a resistance box (Fig. 8-27).

Resistance boxes are of the plug type and the dial type. In the former, the resistors are switched in and out by positioning plugs in respective holes. In the latter, this is done by means of a dial switch.

An assembly of five precision resistors, each resistor having 10 times the value of the preceding resistor, is called a five-section decade box (Fig. 8-27). This box can be set to any desired value from 0 to 9 ohms in steps of 1 ohm. Similar decade boxes are made for resistances of 0.9, 9, 90, 900, 9000 and more ohms.

A resistance bridge (Fig. 8-28) has three arms in the form of resistance boxes r_1, r_2 and r which, together with unknown resistance r_x, form a closed mesh $ADBC$. The junctions C and D are connected to a power source, and the junctions A and B to a galvanometer.

The resistances of r_1, r_2 and r are adjusted until the galvanometer shows a zero deflection. This is an indication that the bridge is at balance, the potentials at points A and B are identical and, as a consequence,

$$V_{CA} = V_{CB} \quad \text{and} \quad V_{AD} = V_{BD}$$

or

$$I_1 r_1 = I_2 r_x \text{ and } I_1 r_2 = I_2 r$$

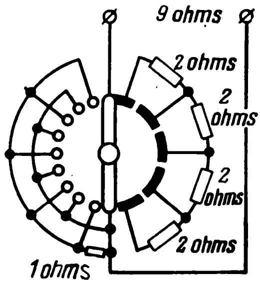


Fig. 8-27. Dial-type five-decade resistance box

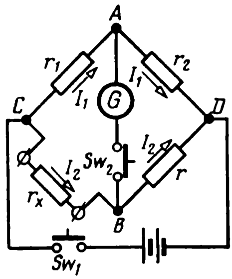


Fig. 8-28. Resistance bridge

Dividing one equality by the other termwise gives

$$I_1r_1/I_1r_2 = I_2r_x/I_2r$$

Hence,

$$r_x = rr_1/r_2 \tag{8-23}$$

As is seen, the unknown resistance is given by Eq. (8-23).

When low-value resistances are measured, an appreciable error is introduced by the resistance of the connecting wires (or flyleads, as they are often called). In such cases, resort is made to the more elaborate double bridges.

If we hold the resistances in three arms of the bridge and the supply voltage at a constant value, indications of the galvanometer will depend only on the value of r_x. Accordingly, the galvanometer scale can be calibrated directly in units of the unknown resistance or in units of the quantity controlling it, say, temperature. Such bridges are called *unbalanced.*

(b) Measurement of Resistance by the Ammeter-Voltmeter Method

The quotient of indications on a voltmeter placed across the unknown resistor, by indications of an ammeter placed in series with that resistor gives its resistance

$$r_x = V/I$$

To maintain the required current in the resistor, a rheostat is usually connected in series with it.

(c) Ohmmeters

Instruments which measure resistances directly are called *ohmmeters* or *megohmmeters*, depending on the value of the resistance being measured.

Ohmmeters are divided into two groups, namely ohmmeters whose indications depend on the supply voltage, and those independent of the supply voltage. Instruments in either group can have two types of measuring circuit, namely a series one and a parallel one.

Ohmmeters in the first group, using a series measuring circuit (Fig. 8-29) are essentially a moving-coil meter movement with a series resistor, r_s, to which is connected the unknown resistor, r_x. The ohmmeter has a supply source of its own—a dry-cell battery.

When the button of the switch Sw is open, the current in the movement is

$$I = c_I\alpha = V/(r_x + r_m + r_s) \qquad (8\text{-}24)$$

where c_I is the current constant of the movement.

It follows from Eq. (8-24) that the deflection of the moving element is

$$\alpha = (V/c_I)\,\frac{1}{r_x + r_m + r_s}$$

The sum $r_m + r_s$ remains constant. So, if the ratio V/c_I also remains constant, the deflection angle α will only depend on the unknown resistance r_x, and the scale of the ohmmeter may then be calibrated directly in units of resistance.

To maintain the ratio V/c_I constant despite variations in the supply voltage, it is necessary to adjust c_I, which is done by varying the magnetic induction in the air gap of the movement by means of a magnetic shunt. A magnetic shunt is an iron bar which is caused to approach or recede

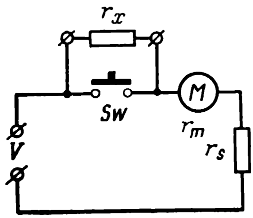


Fig. 8-29. Series-circuit ohmmeter whose indications depend on supply voltage

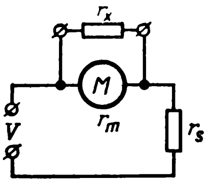


Fig. 8-30. Parallel-circuit ohmmeter whose indications depend on supply voltage

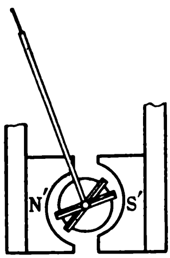


Fig. 8-31. Movement of a ratiometer

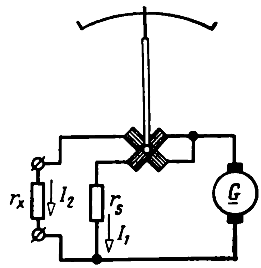


Fig. 8-32. Circuit of a ratiometer-type ohmmeter

from the pole-pieces N' and S' of the movement by a screw (see Fig. 8-1).

After the dry-cell battery and the unknown resistor r_x have been switched in circuit, c_I is adjusted by pressing the button of the switch Sw and varying the position of the magnetic shunt until the ohmmeter reads zero. Then the button is released, and the sought value is read off the scale.

Figure 8-30 shows the same ohmmeter using a parallel measuring circuit in which the unknown resistance r_x is connected in parallel with the movement. It can be shown that if r_m, r_s and V/c_I are held constant, the deflection of the moving element will uniquely depend on the unknown resistance.

Ohmmeters in the second group are essentially a moving-coil meter movement with two coils on a common shaft (Fig. 8-31). Current is conveyed to the coils by ribbon conductors (ligaments) which do not develop any opposing torque.

The currents in the coils flow in opposite directions, so their interaction with the field of the magnet produces two

torques acting likewise in opposite directions. The difference between the two torques causes the moving element to turn through an angle at which the two torques cancel out. The deflection of the moving element is decided by the ratio of the currents in the coils, that is,

$$\alpha = f\,(I_1/I_2)$$

Instruments in which the deflection of the moving element is a function of the current ratio are called *ratiometers*.

One parallel path of a ratiometer-type ohmmeter (Fig. 8-32) is made up of a coil and the unknown resistance r_x; the other parallel path contains the other coil and the series resistor r_s. Recalling that the currents in parallel paths are distributed in inverse proportion to their resistances, we may write

$$\alpha = f\,(I_1/I_2) = f\,(r_x/r_s)$$

Since r_s is constant, the deflection angle is solely dependent on the unknown resistance.

The supply source is usually a generator whose armature is rotated in a permanent-magnet field. The generator is built into the ohmmeter case and cranked by hand.

(d) Measurement of Insulation Resistance

The insulation of electrical installations deteriorates fairly easily, so it is essential to check the insulation resistance at regular intervals.

According to the codes for operation of electrical installations in force in the Soviet Union:

(a) the insulation resistance of lighting and power circuits must be tested with a megohmmeter for a test voltage of 1000 V;

(b) the minimum insulation resistance is 0.5 megohm;

(c) for an insulation test, all fuses are removed (or the respective circuit breakers, cutouts or relays are de-energized), and the insulation resistance is measured between adjacent fuses or past the last fuse (or any other protective device), between any wire and ground, and also between any two wires.

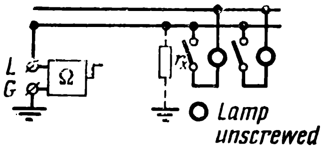


Fig. 8-33. Set-up to measure the insulation resistance of a wire to ground

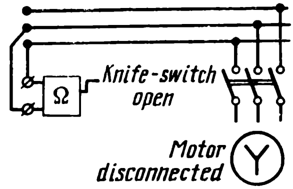


Fig. 8-34. Set-up to measure the insulation resistance between wires

When measuring the insulation resistance of a network in the de-energized condition, one terminal of a megohmmeter, labelled L, is connected to the tested wire, and the other terminal marked G (for ground) is connected to ground (Fig. 8-33). The megohmmeter is then cranked at its nominal rpm, and the insulation resistance is read off the scale.

Then the L-terminal of the megohmmeter is connected to the second wire, and its insulation resistance relative to ground is determined as already explained. To measure the insulation resistance between two wires, the L- and G-terminals are connected each to one of the wires (Fig. 8-34). This procedure also applies to measuring the insulation resistance of electrical machines and apparatus.

8-8. Measurement of Nonelectrical Quantities by Electrical Methods

A wide range of techniques has been developed and has come to be used for the measurement of nonelectrical quantities by electrical means. This has been so, because nonelectrical quantities can then be measured, from a distance, to a high level of accuracy and sensitivity, and, what is more important, continuously.

In most cases, the measurement of a nonelectrical quantity reduces to converting it into an electrical quantity which depends on the measurand in a univalued manner. Then, the resultant electrical quantity is measured and the original nonelectrical quantity is recovered from an appropriate relationship.

The conversion of the unknown nonelectrical quantity to an electrical one is done by devices called transducers*.

All transducers may be divided into two general classes, namely modulating and self-generating transducers. In the former, nonelectrical quantities are converted to variations in circuit parameters, such as R, L and/or C. In the latter, the conversion generates an emf.

The most commonly used modulating transducers are:

(1) **Potentiometric transducers.** These are variable-resistance transducers which depend for their operation on variations in the resistance of a potentiometer (or rheostat) whose wiper is caused to move by variations in the nonelectrical quantity being converted (which may be the level of a liquid, the linear displacement of a part, etc.).

(2) **Strain-gauge transducers.** These, too, are variable-resistance transducers. But in them, the resistance of a wire or semiconductor is caused to vary by strains (deformations).

(3) **Bolometric transducers.** This is a third subclass of variable-resistance transducers, in which variations in the resistance are caused by temperature variations.

(4) **Variable-inductance transducers.** In this type of transducer, the inductance is caused to vary under the action of variations in the position of some of its elements. These transducers are used in the measurement of force, pressure, and linear displacement.

(5) **Variable-capacitance transducers.** In these transducers the capacitance is made to vary by variations in the unknown nonelectrical quantity, such as force, pressure, linear or angular displacement, moisture content, etc.

(6) **Photoelectric transducers** (see Chapter 17). Such transducers generate a photocurrent or photocurrent pulses whose magnitude or frequency is a function of the unknown quantity. They are used in the measurement of illuminance, temperature, transparency or turbidity of liquids, linear dimensions and other physical quantities.

The most commonly used self-generating transducers are:

---

* In fact, "transducer" is a very general term. In many cases, designers and makers may choose it more suitable to call their products "measuring element", "sensors", "pick-ups", etc.— *Translator's note.*

(1) **Induction transducers.** Basically, they convert the unknown nonelectrical quantity into an induced emf. They are used in the measurement of velocity, linear or angular displacement.

(2) **Thermo-emf transducers.** These transducers generate a thermo-emf bearing a specific relation to the unknown nonelectrical quantity—temperature.

(3) **Piezoelectric transducers.** These transducers are based on the property of some crystals to produce an electric potential difference when they are elastically deformed by an applied force, pressure, or other mechanical factors.

(4) **Photoelectronic transducers.** The description of these will appear in Chapter 17.

Any transducer is in effect only a part (although a very essential one) of a rather complex set-up incorporating the transducer proper, connecting wires, a movement with a dial calibrated in units of the unknown quantity, a power source, a stabilizer, rectifiers, amplifiers, etc.

Now we shall discuss the use of some of the transducers in more detail.

(a) Potentiometric Transducers

A potentiometric transducer is essentially a potentiometer whose wiper is caused to move under the action of variations in the unknown nonelectrical quantity x.

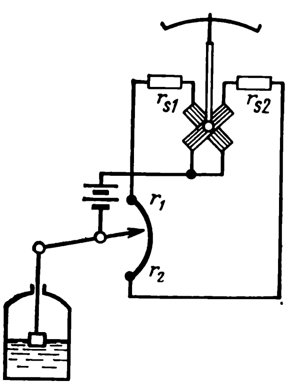


Fig. 8-35. Circuit of a liquid gauge

An example of such a device, used to measure the level of liquids, is shown in Fig. 8-35. The position of the float depends on the level of the liquid. Any change in the latter causes the float to move up or down. Because the float is mechanically linked to the potentiometer wiper, the latter moves, too, and causes a change in the resistances r_1 and r_2, connected in series with the coils of a ratiometer. The change in the current ratio of the coils brings about a change in indications of the instrument which, in this particular case, is called a liquid gauge.

(b) Induction Transducers

An example of this subclass is an induction tachometer which measures rotational speed by converting it to a proportional emf.

A tachometer is a small generator whose armature rotates in the magnetic field of a permanent magnet (Fig. 8-36) and whose emf is proportional to its rpm. The armature is mechanically linked to the shaft of the machine whose speed is to be measured, so the voltmeter connected across the armature of the tachometer is calibrated to read the rpm of that machine.

In an induction tachometer using a permanent magnet NS (Fig. 8-37), the latter is mechanically linked to the shaft of the machine whose speed is being measured. As it rotates, eddy currents are induced in an aluminium disc, *1*, mounted on the common shaft with a pointer, *2*. The interaction of the eddy currents with the magnetic field established by the permanent magnet produces a torque which drives the disc

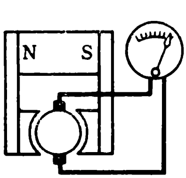


Fig. 8-36. Circuit of an induction tachometer

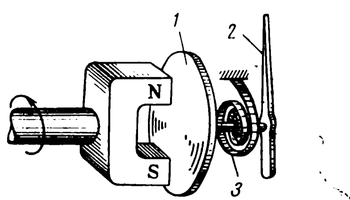


Fig. 8-37. Construction of a tachometer with a revolving magnetic field

and pointer through an angle where the driving torque is balanced by the opposing or control torque supplied by a spring, 3. The tachometer dial is marked, each mark representing a particular rotational speed.

(c) Thermo-EMF Transducers

An elementary thermo-emf transducer is a thermocouple. When combined with a moving-coil meter movement, such a thermocouple makes a thermo-emf thermometer (Fig. 8-38).

As the hot junction, *a*, of thermocouple is heated, a thermo-emf is induced, giving rise to a current flow through the meter movement. As a result, the moving element deflects and the pointer indicates the temperature measured. The wires of the thermocouple must be sufficiently long, so that the cold junction, *b*, has the temperature at which the instrument was originally calibrated.

Most often thermocouples are fabricated from copper against constantan (for temperatures not over 300°C), copper against copel (for temperatures not over 600°C), iron against copel (for temperatures not over 800°C), chromel against copel (for temperatures not over 900°C), chromel against alumel (for temperatures not over 1300°C), platinum against platinum-rhodium alloy (for temperatures up to 1600°C).

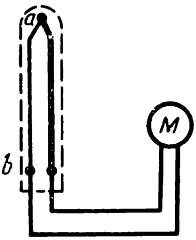


Fig. 8-38. Circuit of a thermo-emf thermometer

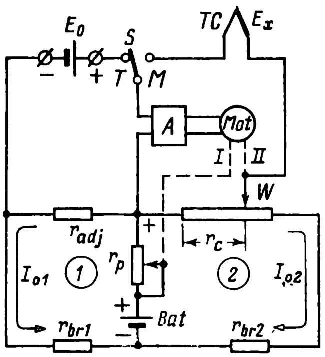


Fig. 8-39. Simplified circuit of an automatic potentiometer

As a protection against mechanical damage and attack by corrosive gases, thermocouples are enclosed in sleeves made of brass, steel, porcelain or some other material.

A thermocouple thermometer is not the only type of instrument using a thermo-emf transducer. For example, wide use is made of automatic potentiometers which can continuously measure and/or record the emf of temperature on a paper chart. A simplified circuit of an automatic potentiometer is shown in Fig. 8-39. The two loops (or meshes), *1* and *2*, of the potentiometer are traversed by operating currents, I_{o1} and I_{o2}. The resistors r_{br1} and r_{br2} make it possible to set the operating currents in a definite ratio. To obtain the desired value for I_{o1}, the switch S is placed in the "T" (test) position. In this position, the difference between the emf of a standard cell, E_0, and the voltage drop $I_{o1}r_{adj}$ that is, the difference voltage $V_d = E_0 - I_{o1}r_{adj}$, is applied, after amplification by an amplifier A, to a reversible motor M, thereby causing it to rotate and move the wiper of the potentiometer r_p. The rotor keeps rotating until I_{o1} reaches the value at which V_d reduces to zero. In this condition, I_{o1} takes on a value required for the measurement. At that instant, the switch S is automatically moved to the "M" (measure) position. At the same time, the motor is disengaged from the wiper of the potentiometer r_p and is engaged with the wiper W of a slide-wire r_c. The difference between the emf E_x generated by a thermocouple TC, and the voltage drop across the slide wire, $I_{o2}r_c$, that is, the difference voltage $V_d = E_x - I_{o2}r_c$, causes the motor M to move the wiper W of the slide-wire until E_x and $I_{o2}r_c$ balance each other. The wiper W of the slide-wire is mechanically coupled to the stylus of a recorder and the pointer of a meter, which display the value of the unknown quantity at balance.

Chapter Nine Transformers

9-1. Purpose of Transformers

A transformer is a static electromagnetic apparatus for converting electric power from a primary a.c. system at one voltage or current to electric power in a secondary a.c. system at a different voltage or current, while keeping the initial frequency unchanged. In other words, a transformer accepts electric power at a voltage V_1 or current I_1 and delivers electric power at a different voltage V_2 or current I_2.

In present-day power systems*, the electricity generated by fuel-fired power stations in localities having an ample supply of coal, petroleum or gas, or by hydro-power stations utilizing the energy of large rivers is transmitted over large distances which can run into thousands of kilometres.

P. N. Yablochkov (1847-1894)

To make long-distance power transmission economical, it is customary to step up the line voltage to tens or even hundreds of kilovolts at the sending end. On the other hand, to make power distribution safe, this voltage has then to be stepped down at the receiving end. It is often that the electricity generated by a power

* A power system is an interconnection of electric power stations by high-voltage power transmission lines.

station is stepped up or down in voltage three or four times before it reaches its users. Obviously, the transformers that do this job must have a very high efficiency.

The first transformer was invented by P. N. Yablochkov, an outstanding Russian scientist and inventor.

9-2. Operating Principle and Design of a Single-Phase Transformer

For its operation, a transformer depends on mutual induction (see Sec. 3-18).

A simplified circuit of a single-phase transformer is shown in Fig. 9-1. Its magnetic circuit, or core, *3*, is assembled from punchings or laminations fabricated from electrical sheet steel carrying 4 or 5% silicon and given a coating of varnish on both sides. The limbs, or legs, of the core give support to transformer windings, *1* and *2*. The supply voltage is impressed on the primary winding, *1*. The resultant power, P_1, is the primary or input power of a transformer. Winding, *2*, is connected to a load, Z, and is called the secondary winding. The power associated with the secondary winding, P_2, is called the secondary or output power.

As a rule, the primary and secondary voltages are different. Accordingly the winding carrying a higher voltage is called the high-voltage (or H. V.) winding, and that carrying the lower voltage is called the low-voltage (or L.V.) winding. Each winding of a single-phase transformer (Fig. 9-1) is made up of two halves put on different limbs of the core and interconnected in such a way that their mmfs are added together to give the total magnetic flux. The greater part of the flux Φ has its path completed through the core, *3*, and is called the *useful flux*, *4*. It links both windings. The part of the magnetic flux which has its path completed through the air gap and links only one winding (*5* or *6*) is called the *leakage flux* of that winding.

In Fig. 9-1, the primary and secondary windings are shown separated to make the drawing to look less crowded. In practice, they are usually arranged concentrically, with the L.V. winding lying closer to the core.

A transformer is a *step-down* one if the primary voltage is greater than the secondary voltage, $V_1 > V_2$; in the oppo-

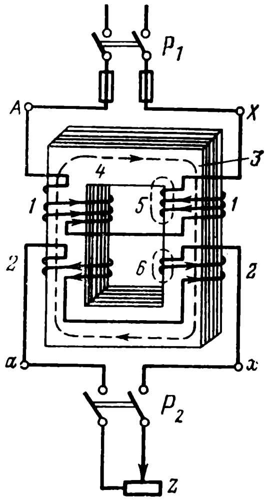


Fig. 9-1. Single-phase transformer

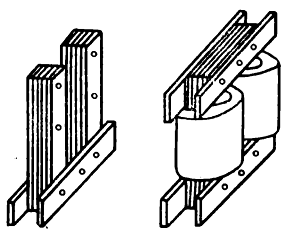


Fig. 9-2. Assembly of a core for a transformer

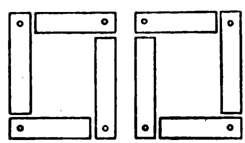


Fig. 9-3. Layout of laminations in a layer of a transformer core during assembly

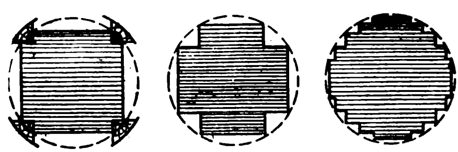


Fig. 9-4. Cross-sections of transformer limbs (legs)

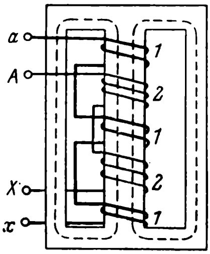


Fig. 9-5. Shell-type transformer

site case, that is, when $V_2 > V_1$, it is called a *step-up* transformer. In a single-phase transformer, the start and finish of the winding on the H.V. side are marked by the letters A and X, respectively; on the L.V. side they are marked by the letters a and x, respectively. In a three-phase transformer, the starts and finishes of the windings are labelled A, B, C and X, Y, Z, respectively, on the H.V. side, and a, b, c and x, y, z on the L.V. side. Assembly of a core for a transformer is illustrated in Fig. 9-2, and the arrangement of individual laminations or punchings in each layer, in Fig. 9-3. The laminations are 0.5 to 0.35 mm thick. The cross-sections of transformer limbs or legs, of various shapes, are illustrated in Fig. 9-4. There exist transformers with branched magnetic circuits (Fig. 9-5).

The ratings of a transformer, that is, power, voltages, currents, and frequency, are stated on its nameplate. Because the efficiency of a transformer is high, the rated power in both windings is assumed to be equal, that is, $S_{1n} = S_{2n}$.

9-3. Performance of a Single-Phase Transformer at No-Load

If we open the knife-switch KS_2 (see Figs. 9-1 and 9-6) and apply the rated voltage V_1 to the primary winding (AX) of a transformer, it will be operating at no-load.

The primary voltage V_1 gives rise in the AX winding to a no-load alternating current, $I_{no-load}$, which does not usually exceed 4 to 10% of the rated current. This current may be deemed to consist of a reactive component $I_{no-load,\ r}$, which sustains the magnetic flux Φ_m, and an active compo-

nent $I_{no-load,\ a}$ which is proportional to no-load loss of power in the transformer

$$I_{no\text{-}load} = \sqrt{I^2_{no\text{-}load,\ a} + I^2_{no\text{-}load,\ r}}$$

Because the no-load current is very small in comparison with the rated current, the power lost in the primary winding as heat, $I^2_{no-load}r$ is usually neglected. This loss is usually termed the iron loss of a transformer, defined as

$$P_{no\text{-}load} = P_i + I^2_{no\text{-}load} r_1 = (\text{approx.})\ P_i$$

The magnetic flux in a transformer is established by the mmf $I_{no-load,\ r}w_1$. Since, however, $I_{no-load,\ a} < 0.1 I_{no-load}$, it is customary to take the mmf of a transformer equal to $I_{no-load}w_1 = F_{no-load}$.

Figure 9-7 shows a vector diagram for a transformer at no-load. The no-load current vector $I_{no-load}$ is laid off in an arbitrary direction, and the maximum value of the pulsating flux is laid off to be in phase with $I_{no-load}$. In the primary and secondary windings, this flux induces emfs [Sec. 6-3, Eq. (6-11)] given by

$$E_1 = 4.44 f w_1 \Phi_m \tag{9-1}$$

and

$$E_2 = 4.44 f w_2 \Phi_m \tag{9-2}$$

These emfs are in quadrature lagging with the magnetic flux. The primary leakage flux, Φ_{1l}, which is in phase with $I_{no-load}$, induces in it an emf of leakage, which is in quadrature lagging with the current:

$$E_{1l} = 4.44 f w_1 \Phi_{1l} \tag{9-3}$$

As has already been shown (see Sec. 6-3),

$$E_{1l} = I_1 \omega L = I_{1,\ no\text{-}load} x_1 \tag{9-4}$$

where x_1 is the inductive reactance of the primary winding due to the leakage flux of that winding.

The voltage drop, $I_{no-load}x_1$, across the primary winding at no-load is less than 0.5% V_1 and may be ignored. Then, by Kirchhoff's second (voltage) law, if we assume that $I_{no-load}r_1 =$ (approx.) 0, the instantaneous voltages equal to the instantaneous emfs will be in anti-phase with each other, that is,

$$v_1 = -e_1$$

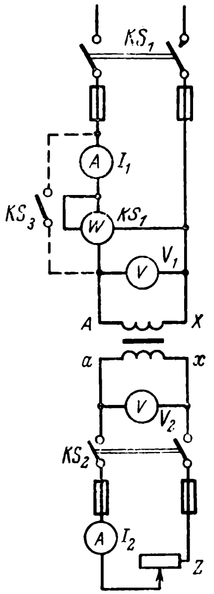


Fig. 9-6. Connection of a transformer in a circuit

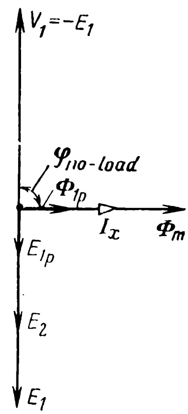


Fig. 9-7. Vector diagram for a transformer at no-load

Since

$$v_1 = V_{1m} \sin \omega t$$

then

$$e_1 = -V_{1m} \sin \omega t = V_{1m} (\sin \omega t + 180°)$$

Thus, the rms voltage and emf

$$V_1 = E_1 = 4.44 f w_1 \Phi_m$$

are equal to, and in anti-phase with, each other (Fig. 9-7), if we neglect E_{1r}.

The current in and the voltage drop across the secondary winding are zero, so the instantaneous voltage and emf, v_2 and e_2, are equal, and so

$$V_2 = E_2 = 4.44 f w_2 \Phi_m$$

The ratio of turns in the windings or the respective emfs is called the *transformation ratio of a transformer*

$$k = E_1/E_2 = 4.44 f w_1 \Phi_m / 4.44 f w_2 \Phi_m = w_1/w_2 \qquad (9\text{-}5)$$

This ratio is usually expressed in terms of the ratio of primary and secondary voltages at no-load, neglecting the voltage drops across the windings, that is,

$$k = V_{no\text{-}load\ 1}/V_{no\text{-}load\ 2} \qquad (9\text{-}6)$$

9-4. Performance of a Single-Phase Transformer under Load and the MMF Diagram

If we close both knife-blade switches, KS_1 and KS_2 (see Figs. 9-1 and 9-6), this will connect a load, z, across the secondary winding, The secondary emf E_2 gives rise in the secondary circuit to a current whose rms value I_2 and direction are, by Lenz's law, such that they tend to maintain the transformer flux Φ_m at a constant value. In other words, *in operation under load, the flux Φ_m is established by the joint action of the mmfs in both windings*

$$\bar{F}_1 + \bar{F}_2 = \bar{F}_{no\text{-}load} \qquad (9\text{-}7)$$

Equation (9-7) shows that the resultant mmf remains practically unchanged and equal to the no-load mmf, $F_{no-load}$. This can be explained as follows. E_1 is proportional to Φ_m ($E_1 \sim \Phi_m$). Also, the voltage drop, $I_1 z_1 < 2$ to 2.5% V_{1n}, is so small as to be negligible. In the circumstances, we may assume that $E_1 \sim V_1$ and $\Phi_m \sim V_{1n}$. Hence, we may approximately deem that the *magnetic flux Φ_m at a constant primary voltage is practically unchanged and remains such in any duty*. The mmf diagram of a loaded transformer is shown in Fig. 9-8.

The magnetic flux Φ_m is in phase with the mmf $F_{no-load}$. The mmf F_2 is shown to be in phase with I_2 which lags behind E_2 by an angle ψ_2. For $F_{no-load}$ to retain its value, the primary winding must produce an mmf given by

$$\bar{F}_1 = \bar{F}_{no\text{-}load} + (-\bar{F}_2)$$

Then, if I_1 at a given instant is flowing from the start to the finish of the winding, I_2 will be flowing from the finish

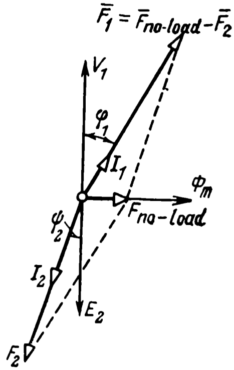


Fig. 9-8. Vector diagram for the mmfs of a transformer

to the start of the secondary winding, and I_1 will automatically rise with increasing I_2. The power factor cos φ_1, which is very small at no-load (cos $\varphi_{no-load} < 0.1$), increases with increasing load owing to the active component of I_2 (φ_1 is less than $\varphi_{no-load}$).

If we ignore the negligible $F_{no-load}$ and assume that $F_1 = F_2$, that is, $I_1 w_1 = I_2 w_2$, then

$$I_1/I_2 = w_2/w_1 = 1/k = E_2/E_1 \tag{9-8}$$

As is seen, the currents in a transformer are inversely proportional to its emfs.

9-5. Variations in Transformer Voltage at Load

As in the case of a generator, it is important to know how the voltage across the terminals of a transformer varies as its load increases from zero to its rated value.

Mathematically, this relationship is written

$$\Delta V\% = \frac{V_{2\,no\text{-}load} - V_{2n}}{V_{2n}} \times 100\% \tag{9-9}$$

and is called the (percent) regulation of a transformer. At $I_2 = I_{2n}$ and cos $\varphi_2 = 1$, the regulation of a transformer is 2 to 3% V_{2n}.

As is seen from the voltage diagram for a loaded transformer at rated currents in the windings (Fig. 9-9, see also

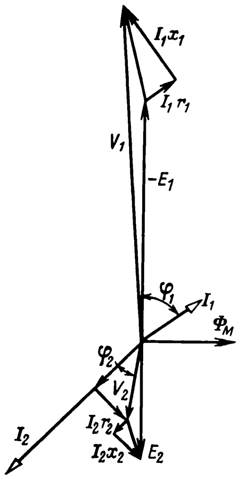


Fig. 9-9. Vector diagram for a loaded transformer

Fig. 6-9 and Sec. 6-4), the primary emf E_1 is smaller than it is at no-load by the voltage drop I_1z_1, because

$$\bar{V}_{1n} = -\bar{E}_1 + (\bar{I}_1z_1 + \bar{I}_1x_1)$$

As E_1 decreases, Φ_m also decreases, and the secondary emf becomes smaller than it is at no-load, that is, $E_2 < < E_{2no\text{-}load}$.

The secondary voltage V_2 of a loaded transformer is deduced by subtracting the voltage drop across the secondary winding from the secondary emf, E_2, of the loaded transformer, and not from the secondary emf at no-load, that is

$$\bar{V}_2 = \bar{E}_2 - (\bar{I}_2r_2 + \bar{I}_2x_2)$$

where x_2 is the inductive reactance due to the leakage flux of the secondary winding. Thus, Eq. (9-9) accounts for the voltage lost in both windings.

9-6. Power Lost in the Windings of a Loaded Transformer

How much power will be lost in the windings of a loaded transformer depends on I_1 and I_2, the winding resistances r_1 and r_2. Mathematically, it is given by

$$P_w = I_1^2r_1 + I_2^2r_2$$

This power is determined by what is known as the short-circuit test. It is carried out, using the test set-up shown in Fig. 9-6. As is seen, the secondary winding is short-circuited, and the primary winding is energized with a small voltage sufficient to give rise to the rated currents, I_{1n} and I_{2n}, in the primary and secondary windings. This voltage is called the impedance voltage, V_i, of a transformer. It is not over 5 to 10% of V_{1n} and is given on the nameplate. In this test, it is immaterial which winding is assumed to be primary.

The supply power measured in this test goes to make up for the power lost in the winding, $P_{w,\,n}$, and the iron loss under short-circuit conditions, $P_{i,\,n}$. The latter is negligible at the low value of the induction B_{sc}. Thus, the total short-circuit power is

$$P_{sc,\,n} = P_{w,\,n} + P_{i,\,n} = (\text{approx.})\ P_{w,\,n} \qquad (9\text{-}10)$$

Hence, the total loss in a loaded transformer at rated current and voltage is

$$\Sigma P = P_{w,\,n} + P_{i,\,no\text{-}load} \qquad (9\text{-}11)$$

9-7. The Three-Phase Transformer

Distribution substations use three-phase transformers (Fig. 9-10).

The core of a three-phase transformer is assembled from electrical-sheet steel laminations as shown in Fig. 9-11. The core has three limbs or legs on which are concentrically arranged the primary and secondary windings of a particular phase, namely AX and ax of the first phase, BY and by of the second, and CZ and cz of the third. For simplicity, Fig. 9-10 shows the windings separated on the respective legs.

The primary and secondary windings can be star or delta connected. Under USSR State Standard GOST 11677-65, three-phase transformers use the star-star connection with the neutral point grounded (symbolized Y/Y_n), the star-delta (Y/Δ) connection, and the star/delta connection, with the neutral point in the star grounded (Y_n/Δ). The numerator lables the connection scheme used on the H.V. side, and the denominator on the L.V. side.

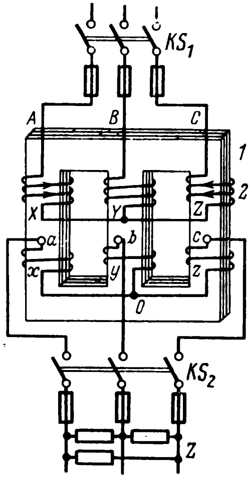


Fig. 9-10. Three-phase transformer

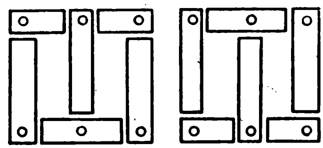


Fig. 9-11. Layout of transformer laminations in assembly

If we consider separately the fluxes established by the individual mmfs, namely $F_A = I_A w_A$, $F_B = I_B w_B$, and $F_C = I_C w_C$, we shall note that the magnetic flux in the AX phase (see Fig. 9-10) has its path completed through the legs B and C, that in the BY phase through the legs A and C, and that in the CZ phase through the legs A and B. When currents simultaneously exist in the three phases, these mmfs are added together as shown in Fig. 9-12.

In addition to the winding connection scheme, the nameplate of a transformer gives the phase-displacement or vector group of the winding connection, such as Y/Y_n-0 or Y/Δ-11. *The phase-displacement or vector group symbol designates the phase displacement between the line emf vector for the L.V. winding relative to the line emf vector for the H.V. winding in the clockwise direction.* The unit of phase displacement is 30°. The 0-phase displacement or vector group is that for which the phase displacement between the line emf vectors is 0°. The 11-group is that for which the phase displacement between the line emf vectors is 330°. This phase displacement is of no consequence to the using of equip-

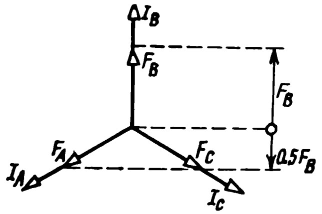


Fig. 9-12. Diagram of mmfs for a three-phase transformer

ment; it is stated so that one can determine whether the transformers involved can be operated in parallel.

In many cases, a supply network at a voltage V_1 must supply two other networks at voltages V_2 and V_3. To meet this requirement, one has to have two transformers with transformation ratios of $k_1 = V_1/V_2$ and $k_2 = V_1/V_3$. Instead of two transformers, use can be made of one transformer having a primary (HV) winding and two secondary windings, of which one is for medium-voltage (MV) and the other for low-voltage (LV). Such transformers are called *three-winding*. For example, in a three-phase transformer with a rated power of $S_n = 6300$ kVA, the primary voltage is 121 kV, the medium voltage is 38.5 kV, and the low voltage is 11 kV.

In a three-winding, three-phase transformer the windings can be connected $Y_n/Y_n/\Delta$-0-11 or $Y_n/\Delta/\Delta$-11-11. The power ratings of the windings are the same. A three-winding transformer is more economical than two separate two-winding transformers.

Multi-winding transformers have a single primary winding and a multiplicity of secondary windings, according to the number of secondary circuits to be energized. Such transformers are, for example, used in radio receivers, TV receiers, tape recorders, etc. Each winding is designed for a particular voltage.

9-8. Control of Transformer Voltage

The output voltage of a generator can be controlled by varying its exciting (field) voltage and, as a consequence, its magnetic flux and emf. In a transformer, the primary vol-

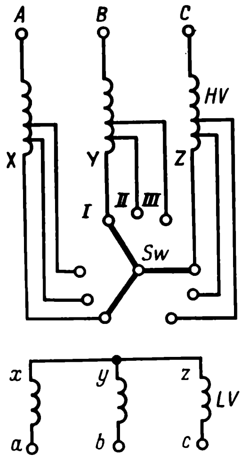


Fig. 9-13. Arrangement of taps on a transformer winding

tage V_{1n} is constant, and so are its magnetic flux Φ_m and emf, therefore the secondary voltage can only be controlled by varying the transformation ratio.

Figure 9-13 shows a three-phase transformer in which each phase on the primary (HV) side has several taps taken to the contacts of a switch *Sw*, mounted on the transformer tank. The handle of the switch is outside the tank lid. The taps are spaced in such a manner that each changes the transformation ratio by $\pm 5\%$. Should the primary voltage drop below its rated value, the switch is moved to another position (position *III* in the diagram) so as to reduce the number of turns and hold the secondary voltage at the rated value. Should the primary voltage rise above its rated value, the switch is moved to position *I* in the diagram, thereby restoring the secondary voltage to its rated value. Such switches are called tap-changers, and they come in two varieties. One variety operates so that it opens the circuit, and this requires that the transformer be de-energized, thereby entailing a disruption in the supply of electricity to consumers. The other variety does its job without interrupting the circuit, and the respective devices are known as under-load tap changers.

9-9. Autotransformers

An autotransformer is a transformer in which some part of the winding is used in both the primary and the secondary circuits at the same time (Fig. 9-14). Autotransformers are advantageous in cases where the transformation ratio is $0.5 \leqslant k \leqslant 2$. They are used to link HV networks operating at 500 and 220 kV, to start induction and synchronous motors, as voltage regulators in a laboratory, etc.

An autotransformer operates as follows. When the primary winding AX is energized by an alternating voltage (see Fig. 9-14*a*), a magnetic flux is established in the transformer core, and an emf E_1 is induced in the winding. The voltage set up across the portion ax which serves as the secondary winding is proportional to the number of turns in that portion. The secondary current I_2 flows in the portion ax, and the primary current I_1 in all of the winding AX. Because the primary and secondary currents flow in opposite directions, the portion ax carries their difference, $I_{ax} = I_1 - I_2$, so this portion may be wound with a finer wire. The autotransformer shown in Fig. 9-14*a* is a step-down one, because $w_1 > w_2$. If we apply V_1 to terminals a and x it will operate as step-up transformer, because then $w_1 < w_2$. Figure 9-14*b* shows a step-down three-phase autotransformer.

Some autotransformers are built so that their transformation ratio can be varied at will (Fig. 9-14*c*). Owing to this feature, they enable the output voltage to be varied from zero to 1.1 V_{1n}. Additional taps on the primary winding make it possible to connect the autotransformer to supply

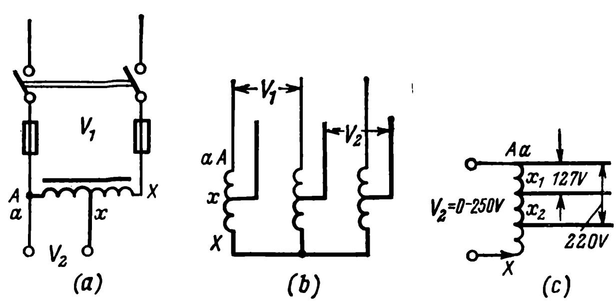


Fig. 9-14. Autotransformers

lines operating at 127 or 220 V. They have a roller contact, terminal X, which rides on the bare (skinned) outer surface of the primary winding and continuously varies the primary voltage in steps of less than 1 V.

A major disadvantage of all autotransformers is that the HV and LV windings are conductively coupled. In the case of a high voltage, the transformation ratio has to be set at 2:1 or 2.5:1, because the insulation resistance of the secondary circuit with respect to ground must be the same as that of the primary circuit.

Among the advantages of autotransformers are the lower weight of copper for the windings, reduced I^2R (heat) losses, and, as a corollary, better efficiency in comparison with conventional transformers.

9-10. Arc Welding Transformers

Conventional transformers cannot be used as welding-arc sources because a prohibitively heavy current is flowing (15 to 20 times the rated value) when an arc is struck and the electrode is touched to the work.

In arc welding transformers, the secondary voltage varies from $V_{2,\, no\text{-}load} = 70$ V at no-load to $V_{2sc} = 0$ at short-circuit, when the electrode is touched to the work.

In the latter case, the short-circuit secondary current, $I_{2,\, sc}$, ought not to exceed the operating current, I_2, by more than 20 to 40%. Such a transformer must have a very drooping external (load) characteristic (Fig. 9-15). If this condition is satisfied, then I_2 will remain practically constant, even though the voltage may fluctuate appreciably owing to variations in the electric-arc resistance, which is essential, if a high-quality arc weld is to be made. To obtain such a characteristic, resort is made to transformers whose windings have appreciable leakage flux Φ_l, or which are fitted with a separate saturable reactor, or have an additional winding on the common core.

In the first case (Fig. 9-16*a*), the primary winding *1* is designed for standard voltages of 220 V or 380 V. The secondary winding, *2*, which is connected in series with a separate saturable reactor, *3*, has a no-load voltage of $V_{2,\, no\text{-}load} = 70$ V, while at the rated secondary current

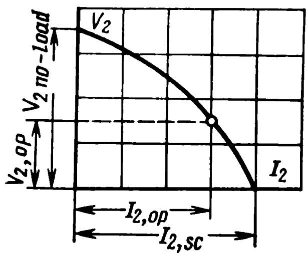


Fig. 9-15. External (load) characteristic of a welding transformer

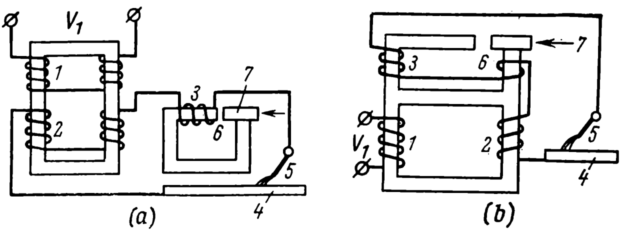


Fig. 9-16. Arc welding transformer

I_{2n} its voltage is $V_2 \approx 30$ V. The welding current between the electrode, *5*, and the work, *4*, can be adjusted by varying the air gap, 6, in the coil, *3*. This is done by shifting the movable part of the core, *7*.

In the second case (Fig. 9-16*b*), the reactor, *3*, and the secondary winding, *2*, are placed on a common magnetic core and are coupled inductively. The movable part of the core, *7*, has the same function as in the arrangement of Fig. 9-16*a*. The efficiency of welding transformers is 83 to 90%, and $\cos \varphi = 0.52$ to 0.62.

9-11. Instrument Transformers

Instrument transformers are used in a.c. circuits in order to extend the range of measurements. Also, such transformers ensure safety to attending personnel, as they isolate instruments, relay coils and other circuit components from high-voltage circuits. The manner in which an ammeter, voltmeter and watthour-meter can be connected via instrument transformers is illustrated in Fig. 9-17.

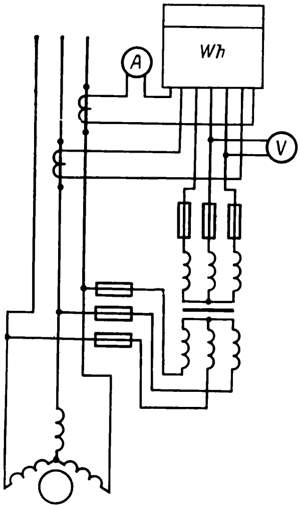


Fig. 9-17. Connection of instrument transformers and instruments

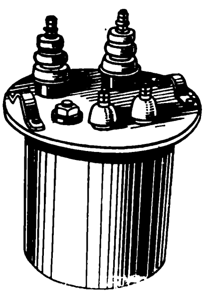


Fig. 9-18. Instrument voltage transformer

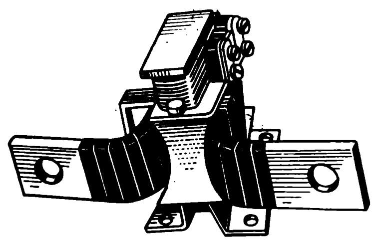


Fig. 9-19. Instrument current transformer

All voltage circuits of the measurement devices are connected to the secondary winding of an instrument voltage transformer whose primary winding is connected to the high-voltage network *ABC*. Its secondary winding is designed for a rated voltage of 100 V. An external appearance of a single-phase instrument voltage transformer is shown in Fig. 9-18. As a protection against inadvertent short-circuits, the primary and secondary circuits of the transformer contain

fuses. The transformation ratio of the transformer, $k_V = V_1/V_2 = w_1/w_2$, may be deemed constant only at rated power. Then,

$$V_1 = k_V V_2 \tag{9-12}$$

can be measured to an acceptable level of accuracy.

The current (series) circuits of ammeters, wattmeters and watthour-meters are connected to the secondary winding of an instrument current transformer. The secondary winding of an instrument current transformer is designed for a rated current of 5A (Figs. 9-17 through 9-19). The primary winding of an instrument current transformer, often consisting of one or two turns of heavy wire, is connected in series with the measured circuit. In this case, the transformation ratio

$$k_I = I_1/I_2 = (\text{approx.})\, w_2/w_1$$

will remain unvarying only when the total resistance of the meter windings and connecting leads does not exceed the limit set for a given transformer. Then,

$$I_1 = k_I I_2 \tag{9-13}$$

No fuses are installed in the secondary circuits of instrument current transformers. The point is that a blown fuse in the secondary circuit would cause the secondary mmf F_2 to disappear, while the primary mmf F_1 would remain as it was before. Under normal operating conditions, these mmfs oppose each other and produce a very small resultant mmf F_x. Obviously, if F_2 were allowed to disappear, this would cause F_x to rise to F_1. Then the magnetic flux in the transformer and the emf in the open-circuited secondary winding would rise to dangerous values (which are likely to cause core overheating, insulation breakdown, and an electric shock to attending personnel).

When an instrument transformer is connected in a high-voltage network, its secondary winding and frame must be grounded.

9-12. Efficiency of a Transformer

The efficiency of a transformer is defined as the ratio of the active or output power, P_2, it delivers, to the active power applied to its input, P_1, or mathematically

$$\eta = (P_2/P_1) \times 100\% = [P_2/(P_2 + P_i + P_c)] \times 100\% \tag{9-14}$$

where P_i is the iron loss found by the no-load test (see Sec. 9-3), and P_c is the copper loss found by the short-circuit test (see Sec. 9-6).

The efficiency of a transformer depends on its load, because the iron loss is constant and the copper loss is proportional to the current squared. If the loading factor is defined as the ratio of power, $S_2/S_n = k_{load}$, then the efficiency of a transformer is given by

$$\eta = P_2/P_1 = \frac{k_{load}S_{2n}\cos\varphi_2}{k_{load}S_{2n}\cos\varphi_2 + P_i + k_{load}^2 P_{c,\,n}} \qquad (9\text{-}15)$$

where $P_{c,\,n}$ is the copper loss at the nominal current, as determined by the short-circuit test.

Calculations and experiments show that transformers have a maximum efficiency when the loading factor k_{load} is approximately from 0.7 to 0.8, because then the copper loss is equal to the iron loss.

9-13. Heat Control in Transformers

When a transformer is operating, its core and windings give up heat. This heat must be abstracted to the surroundings since otherwise the temperature of the transformer might rise to a dangerously high value. The maximum allowable temperature (under Soviet safety codes) is set at 105°C for the windings, 110°C at the surface of the core, and at 95°C for the top layers of cooling oil at an ambient temperature of 35°C*.

As a rule, power transformers are oil-cooled, in which case the windings are cooled well and protected against exposure to moisture and atmospheric oxygen. A power transformer is enclosed in a steel tank filled with mineral oil. For transformers rated at 20 to 30 kVA, the tank has plain walls; for larger transformers the tank has a system of tubes which increase the cooling surface (Fig. 9-20). The top lid of the tank carries bushing insulators for the winding leads.

---

* In the US and UK literature of the subject, this condition is usually stated as "temperature rise", that is, the number of degrees above ambient temperature. The respective figures would then be 70°, 75°, and 60°C.— *Translator's note.*

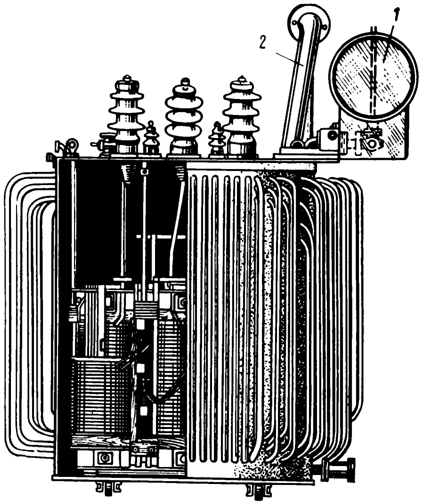


Fig. 9-20. Power transformer

At a power rating of over 100 kVA or even at a lower power rating if the rated voltage is over 6.3 kV, the tank must have an oil conservator, *1*, (also known as an oil expansion tank) connected to the tank by a tube (see Fig. 9-20). On heating, the oil level in the conservator rises; on cooling, it lowers. The capacity of the conservator must be such that it will ensure an ample supply of oil in the main tank at any load and any variations in ambient temperature from —35°C to +35°C. The oil level is indicated by an oil gauge.

At $S > 1000$ kVA, power transformers are fitted with a breather, *2*, connected to the tank and closed by a glass burst diaphragm on the outside. In the case of a fault or breakdown in the transformer, the gases liberated by evaporating oil break the diaphragm, thereby protecting the main tank from being burst instead.

Chapter Ten Alternating-Current Electrical Machines

10-1. Purpose of Alternating-Current Machines. Induction Motors

As a rule, electric energy is generated, transmitted and utilized as a three-phase system of voltages. At a power station, mechanical power is converted into electricity by synchronous generators. At the using end, electricity is converted back to mechanical energy mainly by induction motors.

The three-phase induction motor was invented by M.O. Dolivo-Dobrovolsky in 1889. His motor depends for its operation on a revolving magnetic field. Being the most commonly used type of motor, his machine has since then remained almost unchanged in construction. Simple, inexpensive and reliable in service, the induction motor has good mechanical characteristics and is manufactured in power sizes from a fraction of watt to thousands of kilowatts at voltages of 127, 220, 380, 500, 660, 3000, 6000 and 10,000 V. It is used to drive machines and mechanisms where the rotational speed need not be maintained precisely constant or controlled. The single-phase type of induction motor is the regular feature of household appliances, such as refrigerators, floor polishers, washing machines, and the like.

10-2. The Revolving Magnetic Field

In most cases, a.c. machines are of the so-called inverted construction. This means that the armature winding remains stationary and the part carrying it is called the stator, while the field winding is made rotating and the part carrying it is called the rotor. In the case of a three-phase system, the stator current produces in the machine a revolving magnetic

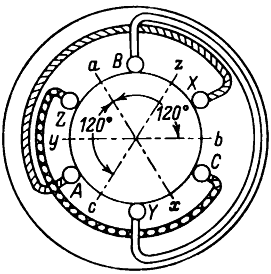


Fig. 10-1. Arrangement of a three-phase winding

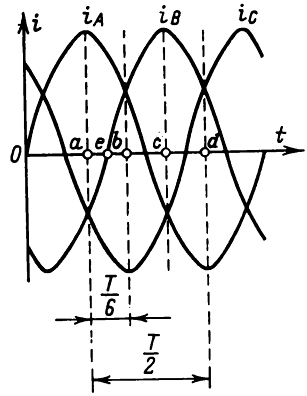


Fig. 10-2. Current waveforms in a three-phase system

field whose speed depends on the frequency f of the applied current and number of pole pair p in the machine. We shall take as an example a stator winding with $2p = 2$ poles, which is energized with a current at 50 Hz and produces a magnetic field revolving at $n = 3000$ rpm.

The three identical stationary windings AX, BY and CZ, arranged on the inner surface of the stator (Figs. 10-1 and 10-8) so that they are spaced 120° apart from one another are traversed by currents making up a three-phase system. The currents, too, are displaced from one another by 120°. The waveforms of the currents i_A, i_B and i_C are shown in Fig. 10-2.

As in Fig. 10-1, each coil in Fig. 10-3 is shown schematically as a single turn, and the coil ends are omitted. Let us assume that the currents flow from the start to the finish of each coil and call this direction *positive*. Then the directions of the currents in each coil in Fig. 10-3 for several instants (a, b and c) will be as shown in Fig. 10-2.

At time a, the current in the coil AX is a maximum and positive (see Fig. 10-2). The currents in the coils BY and CZ are negative, equal to each other, and each is half as great as the current in the coil AX. Thus, in Fig. 10-3a, the current at the start of the coil (terminal A) is flowing away from the reader, while at the starts B and C of the coils BY

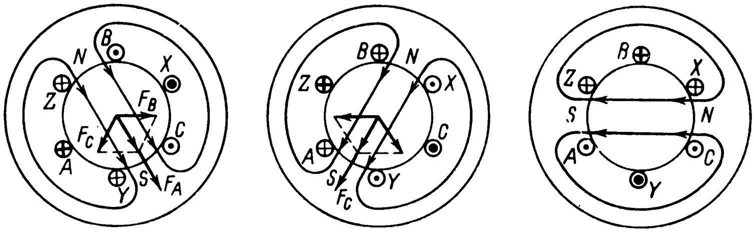


Fig. 10-3. Combination of mmfs at times a, b and c (see Fig. 10-2)

and CZ it is flowing towards the reader. At times b and c, the currents are plotted in a similar way (Figs. 10-2 and 10-3b and c).

It is easy to see that the magnetic lines of force which link the currents flowing in the same direction at times a, b and c are displaced from one another in space in the clockwise direction so that the field revolves through 60° during one-sixth of a cycle and completes a revolution during a cycle.

In Fig. 10-3 it is also seen that when the current in a coil is a maximum, the direction of the revolving field is along the axis of that coil.

Once we know the sequence in which the currents in the coils pass through their peak, or amplitude, values (A, B, C, A, etc. in Fig. 10-2), we can determine the direction in which the field is revolving. If we want to reverse the direction, we must reverse the sequence of current amplitudes in the coils, this is done by interchanging any two wires of the three that connect the coil to the supply mains. Reversal of induction motors is accomplished in a similar manner.

Figure 10-3 shows a two-pole field. If we make the winding so that each phase consists of two series-connected coils, and not of one coil as shown in Fig. 10-1, the field will complete one-half of a revolution during a cycle. Thus, in the general case, the rpm of the revolving magnetic field is

$$n_1 = f_1 60/p \tag{10-1}$$

This is known as the *synchronous* speed.

From examination of Fig. 10-3, it can be seen that the mmfs of the coils F_A, F_B and F_C are combined vectorially. Then, noting the values of the currents i_A, i_B and i_C in the coils AX, BY and CZ at times a, b, and c (see Fig. 10-2), we can show that the total mmf of a three-phase winding is

$$\bar{F}=\bar{F}_A+\bar{F}_B+\bar{F}_C=1.5F_{A,\,\max}$$

Since the maximum mmfs of the coils are the same, then $F = 1.5\ F_{tot}$ remains unchanged throughout each revolution, so $\Phi_{phase} \sim 1.5\ F_{phase} =$ constant.

Each phase of the stator winding links a magnetic flux which, owing to the rotation of the magnetic field, is continually varying in time from $\Phi = 0$ to $\Phi = \Phi_m$. This total flux, which is 1.5 times the amplitude of the pulsating current in each phase, induces emfs e_1 and e_2 in the stator and rotor windings, respectively.

10-3. The Stator Winding of an Induction Motor

The stator winding of an induction motor is more elaborate than is shown in Fig. 10-1. Each phase of a three-phase winding consists of several coils similar to those of the armature winding in a d.c. machine (see Fig. 4-9). The stator leads are labelled *S1*, *S2* and *S3* at their starts, and *S4*, *S5* and *S6* at their finishes.

Figure 10-4 shows a coil having four turns, which will occupy two slots in the stator. The same four turns can be divided into two coils as shown in Fig. 10-5. They are connected in series so that their emfs will be added together. All conductors in a coil are insulated together, so in our further discussion each coil will be shown to have a single turn, irrespective of the actual number of turns in it (Fig. 10-6).

The active coil sides may be arranged in slots in one layer, as shown in Fig. 10-1, or, more often, in two layers, as in the armature of a d.c. machine (see Figs. 4-8 and 4-10).

Let us show how one can determine the number of slots in the stator for a three-phase motor. If the machine has $2p$ poles and m phases, then the number of slots that each phase will occupy per pole will be q (where q may be equal to 1,

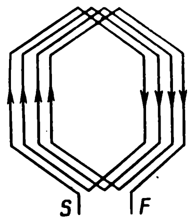


Fig. 10-4. Coil of a stator winding

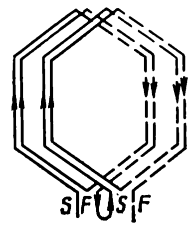


Fig. 10-5. Connection of two coils

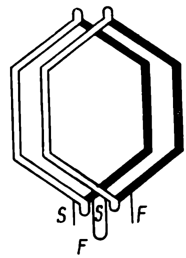


Fig. 10-6. Coil designation

2, 3, 4, 5), usually specified in advance when designing a machine.

Then the total number of stator slots will be

$$Z = 2pmq \tag{10-2}$$

Let $2p = 2$, $m = 3$ and $q = 2$. Then

$$Z = 2 \times 3 \times 2 = 12$$

If we use a two-layer winding, the number of coils will also be 12. Such a winding is shown in Fig. 10-7. Each phase occupies $Z/3 = 12 \div 3 = 4$ slots divided between two coils arranged to lie within the sphere of action of the unlike poles, that is, two pole pitches, τ, apart. The pole pitch in electrical degrees is always equal to 180°.

Slots are distributed among the phases in the following manner. Because $q = 2$, we may arbitrarily take it that at the first pole pitch the phase S1S4 occupies slots *1* and *2*. At the second pole pitch the same phase occupies slots *7* and *8*, because $\tau = Z/2p = 12 \div 2 = 6$ teeth. The phase S2S5 is displaced in space through 120° or through $2/3\tau$, that is, by $6 \times 2/3 = 4$ teeth and occupies slots *5*, *6* and *11*, *12*. The slots are marked on the upper layer of the active sides. Obviously, the phase S3S6 occupies the remaining slots—*8*, *9* and *3*, *4*. For the emfs of each phase to be added together, the sides in each coil are connected in series aiding—the finish of the first to the start of the second, and so on, and the coils are connected in series opposition—the finish of the first to the finish of the second (Fig. 10-7).

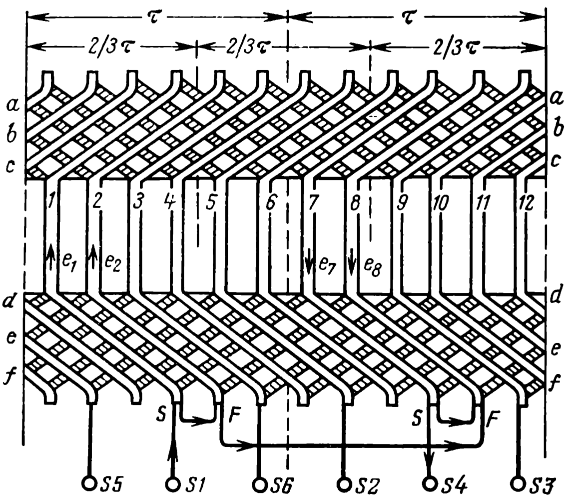


Fig. 10-7. Developed view of a two-layer winding

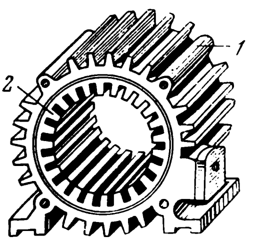


Fig. 10-8. Stator of an induction motor less its winding

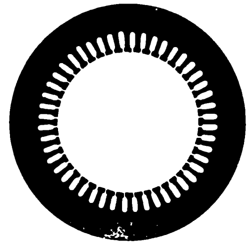


Fig. 10-9. Punching for a stator core

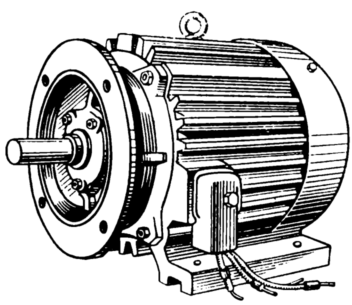


Fig. 10-10. Three-phase squirrel-cage induction motor

Then, for example,

$$e_{S1} = e_1 + e_2 - (-e_7 - e_8)$$
$$= e_1 + e_2 + e_7 + e_8$$

So that the winding can be connected to a three-phase supply line, it is connected in a star or a delta.

The stator of an induction motor, less its winding, is shown in Fig. 10-8. It consists of an outer cast-iron, aluminium or steel frame *1*, into which is press-fitted core *2*, assembled from electrical-sheet-steel laminations (Fig. 10-9). The laminations are insulated from one another by a coating of varnish. In enclosed motors, the outer finned surface of the stator is forced-air-cooled by a blower. A fully assembled motor is shown in Fig. 10-10.

10-4. The Rotor Winding of an Induction Motor

The *rotor* (Fig. 10.11), or the revolving part of an induction motor, is placed inside its stator (Fig. 10-8). It is a cylinder assembled from electrical-sheet-steel laminations in about the same manner as the stator, and has slots on its outer surface. The slots accept copper bars short-circuited at the ends by copper rings to form the rotor winding. In this case, the slots are circular in section, and the winding shown separately in Fig. 10-12 is known as a squirrel cage. In fact, the slots may be given any other shape, and the short-circuited squirrel-cage winding can be obtained by pouring molten aluminium into the slots; at the same time, the shorting end rings are cast, complete with fan blades. Such machines are called squirrel-cage-rotor motors (see Fig. 10-10). Their rotor winding is a *polyphase* one.

Instead of a squirrel-cage structure, the rotor (Fig. 10-13) may have a winding similar to the stator winding (see Fig. 10-7). This type is known as the phase-wound rotor; it is shown in Fig. 10-14. In this case, the three leads, *R1*, *R2* and *R3*, from the winding in slots *1* are connected to three slip-rings *2* mounted on a shaft *3* and insulated from one another and from the shaft. Brushes riding on the slip-rings connect the rotor winding to a rheostat which is used to

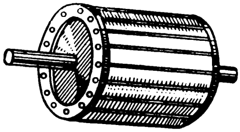


Fig. 10-11. Squirrel-cage rotor of an induction motor

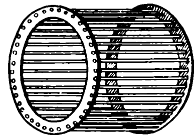


Fig. 10-12. Squirrel-cage structure

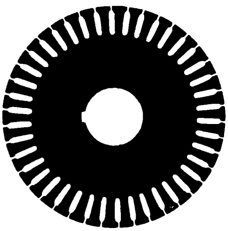


Fig. 10-13. Punching for a rotor core

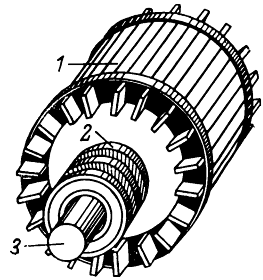


Fig. 10-14. Phase-wound rotor of an induction motor

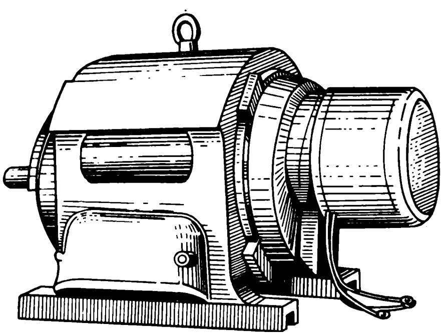


Fig. 10-15. Three-phase phase-wound-rotor (slip-ring) induction motor

start the motor or to regulate its speed. This is a phase-wound-rotor (or slip-ring) motor; it is shown in Fig. 10-15.

10-5. The Operating Principle of an Induction Motor

The emfs induced in the rotor by the revolving magnetic field give rise in its short-circuited conductors to the secondary currents i_2 which interact with the revolving magnetic field of the stator. The conductors in the rotor are acted upon by electromagnetic forces which are tangent to the surface of the rotor (the right-hand rule). On combining, the electromagnetic forces and their moments produce a resultant electromagnetic torque at the rotor shaft, which drives the rotor in the same direction as the stator field is revolving.

The rotor rpm, n_2, must be lower than that of the stator, n_1, because it is only then that the stator field will move relative to the rotor conductors travelling in the same direction and will induce in them the secondary currents, i_2, essential for the operation of the motor.

Such a speed is called *asynchronous* (and, in fact, such motors are sometimes referred to as asynchronous machines). The quotient of the difference between the synchronous speed (that of the stator field) and the actual speed of the rotor to the synchronous speed, expressed as a ratio or a percentage, is called *slip*

$$s = (n_1 - n_2)/n_1$$

or

$$\text{per cent } s = \frac{n_1 - n_2}{n_1} \times 100\% \qquad (10\text{-}3)$$

The slip of an induction motor can vary from 1, or 100%, when the rotor is stationary, to zero when the rotor is rotating at the synchronous speed (that of the stator field). As the load on the shaft increases, s also increases, because it is only then that E_2, I_2 and, as a consequence, the rotor torque can be sufficiently large. The rated slip for induction motors is from 1 to 6%. At no-load ($P_2 = 0$), the slip is practically zero.

At no-load, the rotor current is relatively small. When some load is applied to the shaft ($P_2 \neq 0$), it increases.

The resultant magnetic flux Φ of the machine is produced by the joint action of the mmfs F_1 and F_2, as in a transformer (see Fig. 9-8, Sec. 9-4). However, F_1 and F_2 can be combined only when they are stationary relative to each other, as in a transformer. Let us show that in an induction motor F_1 and F_2 are stationary relative to each other, although they are revolving in space at n_1, that is, at synchronous speed.

The frequency of the stator current, f_1, is proportional to n_1, and that of the rotor current, f_2, is proportional to $(n_1 - n_2)$. Also,

$$f_2 = p\,(n_1 - n_2)/60 = pn_1s/60 = f_1s \tag{10-4}$$

When the rotor is stationary, $f_2 = f_1 \times 1 = f_1$. At synchronous speed, $f_2 = f_1 \times 0 = 0$. In operation at full load, $I_1 = I_{1n}$, when, say, $s_n = 2$ to 4%,

$$f_2 = f_1s = 50 \times 0.02 \text{ to } 50 \times 0.04$$

that is, 1 or 2 Hz.

The stator current I_1 produces in the stator winding an mmf, F_1 revolving at synchronous speed n_1, while the rotor current I_2 produces in the rotor winding an mmf which revolves relative to the rotor at $n_3 = f_2 \times 60/p$. The rotor itself is rotating in the same direction at n_2. Therefore,

$$n_2 + n_3 = n_2 + f_2 \times 60/p = n_2 + f_1s \times 60/p$$

$$= n_2 + n_1s = n_2 + n_1\,\frac{n_1 - n_2}{n_1} = n_1$$

Thus, an induction motor may be treated as a transformer with a rotating secondary winding, to which we may apply all the reasoning developed in Sec. 9-4, and we shall do so in our further discussion.

It is to be noted that in a transformer (see Fig. 9-7), the no-load current is $I_{no\text{-}load} = 4$ to 10% of I_n. In an induction motor, because there is an air gap between the stator and the rotor, the no-load current is $I_{no\text{-}load} = 20$ to 40% of I_{1n}.

10-6. EMFs in the Stator and Rotor Windings of an Induction Motor

When the rotor is stationary, the revolving magnetic field induces in each phase of the stator and rotor windings the following emfs (see Sec. 9-3):

$$E_1 = 4.44 f_1 w_1 \Phi k_{w1} \tag{10-5}$$

$$E_2 = 4.44 f_1 w_2 \Phi k_{w2} \tag{10-6}$$

where k_{w1} and k_{w2} are the winding coefficients accounting for the design of the windings.

When the motor is running, the rotor speed n_2 and slip s vary, according to the load applied to its shaft (the retarding or load torque).

Accordingly, the frequency of the rotor emf (current), $f_2 = f_1 s$, may take on various values and, as a consequence, the emf in each phase of the rotating rotor will vary in direct proportion to slip:

$$E_{2s} = 4.44 f_2 w_2 \Phi k_{w2}$$

or

$$E_{2s} = 4.44 f_1 w_2 \Phi k_{w2} s \tag{10-7}$$

It is customary to express the emf of a revolving rotor in terms of the emf of the same rotor when it is stationary. This can be done, if we take their ratio

$$E_{2s}/E_2 = f_2/f_1 = f_1 s/f_1 = s$$

Then,

$$E_{2s} = E_2 s \tag{10-8}$$

As a consequence, the emf of a rotor varies over a wide range in the course of operation of the motor. At $s = 1$, $E_{2s} = E_2$, whereas at $s = 0$, it is $E_2 \times 0 = 0$.

10-7. The Impedance of the Rotor Winding

In operation, in addition to the common magnetic flux which links both windings, there also exist separate fluxes called the leakage fluxes, Φ_l (Fig. 10-16). They are respon-

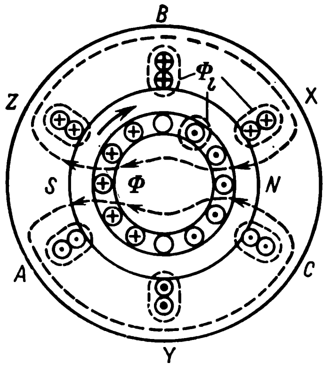


Fig. 10-16. Operation of an induction motor with an active current in its rotor

sible for the reactive impedances x_1 and x_2. When the rotor is stationary, x_2 and x_1 are found as for a transformer (see Sec. 9-3), that is

$$x_2 = \omega L_2 = 2\pi f_1 L_2$$

When the rotor is revolving,

$$x_{2s} = 2\pi f_1 s L_2$$

Therefore, if we take the ratio

$$x_{2s}/x_2 = f_1 s / f_1 = s$$

we can express the reactive impedance of a revolving rotor in terms of the same impedance when the rotor is stationary

$$x_{2s} = x_2 s \qquad (10\text{-}9)$$

As the load on the machine varies, x_{2s} is continually varying, too, from

$$x_{2s} = x_2 s = x_2 \times 1 = x_2$$

to

$$x_{2s} = x_2 s = x_2 \times 0 = 0$$

In normal-type motors, at f_2 varying from 0 to 50 Hz the resistance r_2 may be deemed constant and independent of the rpm (see Sec. 10-10).

10-8. Currents in the Rotor Winding

The current in a phase winding of the rotor is decided by the emf and impedance of the rotor

$$I_2 = E_{2s}/\sqrt{r_2^2 + x_{2s}^2} \tag{10-10}$$

As the rpm is varied, the current also varies in magnitude as a function of E_{2s} and $z_2 = \sqrt{r_2^2 + x_{2s}^2}$, and in phase as a function of the relative values of r_2 and x_{2s}. At starting, so long as the rotor is stationary, I_2 is a maximum, because at $s = 1$, the emf $E_{2s} = E_2 s = E_2$ is also a maximum. The phase difference between the current and E_2 is likewise a maximum, because x_2 = 8 to 10 times the rotor resistance r_2. As soon as the rotor starts revolving, the current and phase difference decrease.

10-9. The Torque of a Motor

The torque of any a.c. motor is governed by its magnetic flux Φ and the active current, namely

$$T = k_T \Phi I_2 \cos \psi_2 \tag{10-11}$$

where k_T is a constant dependent on the motor design.

Figure 10-17 shows how a squirrel-cage induction motor can be connected to its load. When the knife-blade switch is closed, the rotor current $I_{2,\,start}$ is at first a maximum, because a stationary rotor has a maximum emf. However, the starting torque is only 50% to 40% of the maximum torque. The point is that at starting x_2 = 8 to 10% of the rotor resistance r_2, and the phase difference ψ_2 between E_2 and $I_{2,start}$ is close to 90°. As a consequence, the active component of the rotor current, $I_{2,\,start} \cos \psi_2$, is small (Fig. 10-18). For present-day induction motors, *the ratio of starting to rated torque* is T_{start}/T_n = 1 to 1.5, with the ratio of starting to rated current, $I_{2,\,start}/I_{2,\,n}$, approximately ranging between 4.6 and 6.5.

After starting, as the motor picks up speed, the slip s and the rotor emf E_{2s} decrease, and this brings about a decrease in I_2. Since, however, the inductive reactance x_{2s} of the rotor also decreases, then at a constant r_2 the phase diffe-

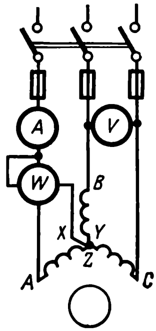


Fig. 10-17. Connection of a squirrel-cage induction motor

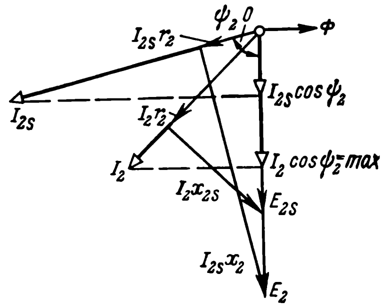


Fig. 10-18. Vector diagram for the rotor circuit

rence ψ_2 will also decrease, while the active component $I_2 \cos \psi_2$ will build up. This signifies that the torque T also builds up. This build-up will continue until x_{2s} becomes equal to r_2. At that instant, the right-angled voltage-drop triangle becomes an equilateral one ($I_2x_{2s} = I_2r_2$) (Fig. 10-18), the active current $I_2 \cos \psi_2$ reaches its peak value, and so does the torque ($T = T_m$). As the rpm, n_2, keeps rising, x_{2s} becomes smaller than r_2, and the latter produces an increasing effect on the current, so that as E_{2s} decreases still more, $I_2 \cos \psi_2$ also decreases, and so does the torque T. The ratio T_m/T_n is usually from 1.8 to 2.5, and is called the *overload capacity* of a motor.

As we have seen, the electromagnetic torque of a motor is a function of slip, $T = f(s)$ at $V_1 =$ constant (Fig. 10-19, curve *1*). A motor develops its rated torque T_n at a rated slip, $s_n = 0.02$ to 0.06. The maximum torque T_m is developed at what is known as critical slip, $s_c =$ (approx.) 0.2. At $s = 1$, a motor develops its starting torque T_{start}.

As will be recalled, the magnetic flux Φ is approximately proportional to the stator voltage V_1 and T is proportional to $\Phi I_2 \cos \psi_2$. So, because $I_2 \cos \psi_2$ is proportional to E_{2s}, Φ, and V_1, we may write

$$T \sim V_1 V_1 = V_1^2 \qquad (10\text{-}12)$$

Or, in words, the *torque of an induction motor is proportional to the square of the voltage applied to the stator winding.* This

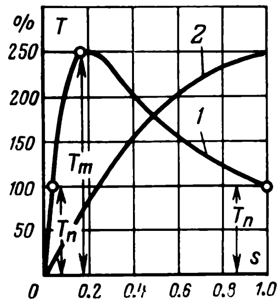


Fig. 10-19. Motor torque as a function of slip

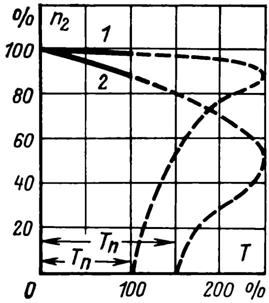


Fig. 10-20. Mechanical characteristics of a motor

relation is essential to the use of induction motors, because a fall in the supply-line voltage to, say, $0.8V_{1,n}$ will inevitably cause the torque to drop from its maximum value to $0.8^2T_m = 0.64T_m$, and the motor will be unable to carry even an insignificant overload and will stop.

The relation $n_2 = f(T)$ at V_1 and f held constant is called the *mechanical characteristic* of a motor. In Fig. 10-20, this characteristic is plotted in coordinates $(n_2/n_1) \times 100\%$ and $(T/T_n) \times 100\%$. Its effective portion between 0 and T_n is shown by the solid line. Curve *1* plotted when the rotor is short-circuited is called the *natural characteristic*. This is a flat characteristic, as in the case of a shunt-wound d.c. motor (see Fig. 4-26).

Curve *2* is called the *artificial characteristic*. It is more drooping than curve *1*, and it is realized when a series resistor is placed in the circuit of a phase-wound rotor. This resistor can be utilized to control the rpm of the motor (as is actually done in crane and hoist drives).

10-10. Starting of Induction Motors

So that the windings of an induction motor can conveniently be connected into a star or a delta at will, their terminals are arranged on a terminal plate as shown in Fig. 10-21. Each phase of the stator is designed for some specific phase voltage V_p. So the stator may be connected to two supply

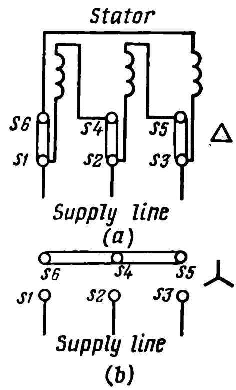


Fig. 10-21. Arrangement of the stator-winding terminals

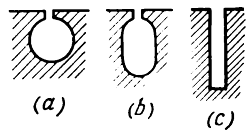


Fig. 10-22. Slots used in induction-motor rotors

lines whose rated voltages differ by a factor of $\sqrt{3}$. For example, if the nameplate voltage of a motor is $V_n = 220$ V for a delta-connected stator (Fig. 10-21*a*), then for a supply voltage of $V_s = 380$ V the stator windings of the same motor must be connected into a star (Fig. 10-21*b*). Another motor may likewise operate from supply lines with $V_s = 380$ V and $V_s = 660$ V, if its windings are suitably connected.

A squirrel-cage induction motor can be started and stopped simply by closing and opening a knife-blade switch (see Fig. 10-17). A disadvantage of this method of starting is that the starting current is rather heavy, being $I_{start} = 4.5$ to 6.5 times the rated current, while the starting torque is relatively small, being $T_{start} = 1$ to 1.5 T_n.

As a way to improve the starting performance of squirrel cage motors, the circular slots in the rotor (Fig. 10-22*a*) are replaced by deeper ones (Fig. 10-22*b*). Then, at starting, when $f_2 = f_1$, the current is forced to flow closer to the surface of the wires, and the resistance of the rotor winding increases. This entails an increase in the active current, $I_{2,start} \cos \psi_2$ (see Fig. 10-18) and, as a consequence, in the initial starting torque, T_{start}. In motors rated at 120 to 150 kW, the slots are made as deep and narrow slits (Fig. 10-22*c*).

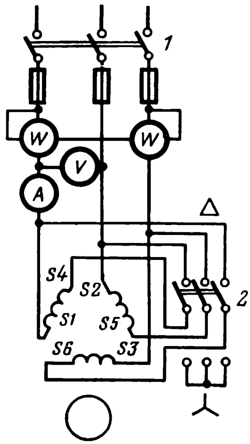


Fig. 10-23. Starting of a motor by re-connecting its stator from a star to a delta

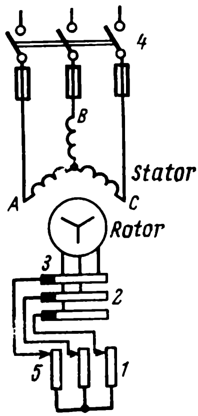


Fig. 10-24. Connection of a slip-ring (phase-wound) induction motor

If the stator windings of a motor are delta-connected, they may temporarily be re-connected into a star in order to reduce the starting current (Fig. 10-23). This is done by moving down the switch *2* before cutting in the knife switch *1*. The motor is set in motion at a phase voltage V_p, reduced by a factor of $\sqrt{3}$, and at a current I_1, reduced by a factor of 3. After starting, the switch *2* is moved upwards, and the motor goes on to operate at $V_1 = V_n$. This form of starting may only be used at no-load, such as in the drive of a fan. The point is that the initial torque in this form of starting is reduced by a factor of $\sqrt{3} \times \sqrt{3} = 3$.

Figure 10-24 shows a starting scheme for a phase-wound induction motor. The contact arms *5* of a rheostat *1* are connected by wires to the brushes that ride the slip-rings *2* of the rotor. To begin with, the rheostat is set to a maximum resistance, the knife switch *4* is then closed, and the rotor is set in motion. The presence of r_{rh} in the rotor circuit serves to bring down the starting current in the rotor and, as a consequence, in the stator to the desired value. As the rotor picks up speed, the current in its winding decreases, and

resistance of the rheostat is gradually reduced to zero. Another advantage of this starting method is that the increased resistance $r_2 + r_{rh} = x_2$ at starting serves to produce a maximum $I_2 \cos\psi_2$ and a maximum torque as shown in Fig. 10-19 (curve *2*). Thus, this scheme may be used to start a motor at full load.

To stop the motor in such a case, the knife switch is opened, and the knob of the starting rheostat is placed to the starting position.

10-11. Speed Control of an Induction Motor

The speed control of an induction motor by a rheostat in the rotor circuit entails heavy losses of power in the rheostat, if the motor is used in continuous duty. Also, this method is only applicable to phase-wound (slip-ring) motors. The principle underlying this method of control may be explained by reference to Fig. 10-25. This figure shows a sketch of a motor and its power diagram where the power P_1 applied to the stator is shown in the form of a flow. The revolving magnetic field transfers to the rotor the electromagnetic power $P_{em} = P_1 - P_i = T\omega_1$, where P_i is the iron loss in the stator, and ω_1 is the synchronous speed in radians ($\omega_1 = 2\pi n_1/60$). The power at the rotor is $P_2 = P_{em} - P_{c,rot} = T\omega_2$, where $P_{c,rot}$ is the copper loss in the rotor winding. If we place a rheostat in the rotor circuit of a motor operating at T = constant, then the power at the rotor will be $P'_2 = P_{em} - (P_{c,rot} + P_{rh}) = T\omega'_2$, where P_{rh} is the power dissipated as heat in the rheostat. Now it is clear that $\omega'_2 < \omega_2$, and the percent reduction in speed is equal to the percent loss in the rheostat, which is wasteful of power.

The speed of squirrel-cage motors can be controlled by varying the supply frequency f_1 or the number of poles p because

$$n_1 = f_1 60/p$$

The frequency f_1 is varied in special-purpose installations only, and is used but seldom. Speed control by variation of the number of poles is utilized in multi- (two-, three- and four-) speed motors. In such motors, the stator windings

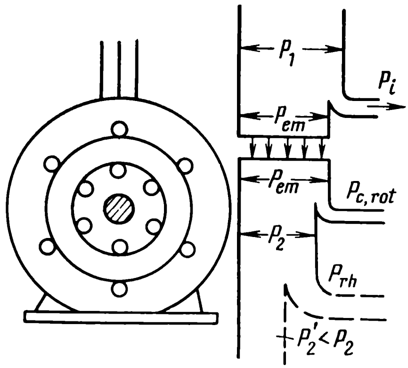


Fig. 10-25. Power diagram of an induction motor

have several identical parts in each phase. These parts can be connected in series or parallel by a switch in such a way that the ensuing redistribution of current changes the number of poles, and this leads to a changed speed, n_1 (3000-1500 and 1000-500 rpm). This arrangement produces a stepped form of speed control. Such motors are used as drives in some machine-tools, as this makes it possible to simplify the speed-change box of the machines.

10-12. The Single-Phase Induction Motor

At power ratings under 0.5 kW, the most commonly used type is the single-phase induction motor. Its connection in circuit is shown in Fig. 10-26*a*. It has a *main* (*running*) *stator winding*, *1*, which is similar to two phases of a star-connected three-phase winding, and a squirrel-cage rotor, *3*. The alternating current I_1 in the main stator winding establishes a pulsating magnetic field which does not produce any starting torque. To produce this torque, the stator has an *auxiliary starting winding*, *2* displaced from the running winding through 90°. This winding is connected in series with a capacitor to the same supply line as the running winding, *1*. The current I_2 in the starting winding is displaced in phase from the current I_1 by a quarter of a cycle.

The two currents with a phase difference of a quarter-cycle and flowing in two windings displaced through 90° from each other produce a two-phase revolving magnetic

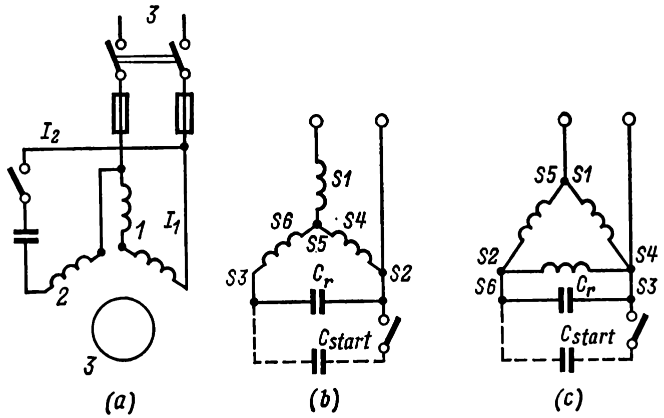


Fig. 10-26. Circuit and connection of a single-phase induction motor

field. In the squirrel-cage rotor, such a field induces emfs and currents which interact with the field and produce a torque. As soon as the rotor is set in motion, the starting winding *2* is de-energized, and the rotor keeps rotating in the pulsating magnetic field established by the running winding, as a single-phase rotor.

How this happens can be explained as follows. Let two mmfs, F_1 and F_2, constant and equal in magnitude (Fig. 10-27*a*) revolve at a speed n_1 in opposite directions, so that they complete one revolution each period (cycle) of the current. When the vectors $\overline{F}_1$ and $\overline{F}_2$ take up the positions shown in Fig. 10-27*a*, their sum is $\overline{F}_1 + \overline{F}_2 = 2\overline{F}_1$. The axis of the resultant mmf is coincident with the axes of the component mmfs.

One-sixth of a cycle later (Fig. 10-27*b*), the mmfs turn through 60° in opposite directions; on combining, they give a resultant mmf equal to $\overline{F} = \overline{F}_1 + \overline{F}_2$. In another quarter of a cycle (Fig. 10-27*c*), their sum is equal to zero, etc. However, the axis of the pulsating mmf remains stationary. Hence, two mmfs equal in amplitude and revolving in opposite directions at an equal speed produce a resultant mmf which pulsates at the frequency of the current along the sta-

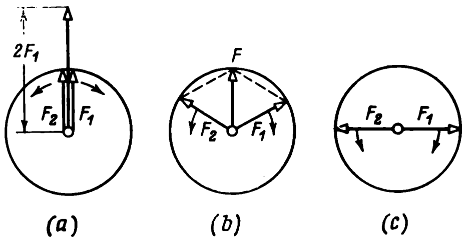


Fig. 10-27. Compensation of a reverse mmf

tionary axis and alternately reaches a positive and a negative peak equal to the arithmetic sum of the two revolving mmfs.

Thus, if a single-phase motor were started with its starting winding *2* open-circuited (Fig. 10-26*a*), the pulsating magnetic field established by the pulsating mmf, F_{start}, of the stator might be deemed to consist of two equal fields revolving in opposite directions at the same speed and each established by an mmf of its own. These fields would induce in the stationary rotor winding two emfs and two equal currents which would interact with their respective fields to produce two equal torques acting in opposite directions. Obviously, the rotor would not rotate under such conditions.

When the starting winding is energized, the two-phase revolving field does produce a torque, and the rotor revolves as it does in a three-phase motor, with the field at a speed n_2 approximating n_1, that is, at a nearly synchronous speed. The field with which the rotor is revolving is called *direct*, and the flux linking the running winding is called the *direct flux*, Φ_d.

If, now, we de-energize the starting winding, when the motor is already running, the rotor, as already noted, will keep running. This happens for the following reason.

The direct flux Φ_d induces an emf, $E_{2,d}$, and a current, $I_{2,d}$ in the rotor revolving at the speed n_2. The result is the production of a direct torque, T_d, similar to that in a three-phase motor. Now, the slip $s_d = \frac{n_1 - n_2}{n_1} =$ (approx.) 0, because n_2 is equal to n_1 very nearly.

The second mmf revolves against the rotor and produces what is known as the *reverse flux* (F_r and Φ_r, respectively).

Relative to the rotor, they revolve at a speed $n_1 + n_2 =$ $=$ (approx.) $2n_1$, that is, with a slip equal to $s = (n_1 + n_2)/n_1 =$ $=$ (approx.) 2. The frequency of $E_{2,rev}$ and $I_{2,rev}$, induced by the reverse stator flux in the rotor is approximately equal to $2f_1$. Accordingly, the reactive impedance of the rotor winding, $x_{2,rev} = 2\pi f_2 L_2$, is so high that $I_{2,rev}$ lags behind $E_{2,rev}$ by almost 90°.

Because of the 90-degree phase difference, almost all of $I_{2,rev}$ is reactive, and its interaction with Φ_{rev} produces a very small torque T_{rev} which opposes the direct torque. The resultant torque is then

$$T = T_d - T_{rev} = \text{(approx.) } T_d$$

and the motor can operate normally.

A similar situation arises in a three-phase motor. When the motor is running and one of the phases is de-energized, the rotor keeps rotating, provided the load does not exceed 50 to 55% of the rated value. However, such a motor, with one of the phases de-energized or open-circuited, will not start.

The reasoning set forth above explains why a three-phase induction motor can operate from a single-phase supply line (see Fig. 10-26*b* and *c*). In this case, as with a three-phase motor, the phase voltage V_p of the stator winding must remain constant in any scheme of connection, that is, $V_p = V_{p,n}$. To meet this requirement, the motor has a capacitor which is permanently connected in the motor circuit and is called the running capacitor, C_r. Its value can approximately be calculated as follows

— for the arrangement of Fig. 10-26*b*

$$C_r = \text{(approx.) } 2800\ I_n/V$$

— for the arrangement of Fig. 10-26*c*

$$C_r = \text{(approx.) } 4800\ I_n/V$$

where I_n is the nameplate phase current and V is the nameplate line voltage. If one desires to obtain a high starting torque, a starting capacitor C_{start} is brought in circuit. Where the motor circuit contains a permanently connected running capacitor C_r, one has what are known as *capacitor motors*.

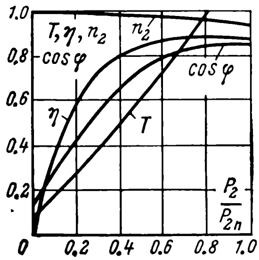


Fig. 10-28. Load characteristics of an induction motor

The maximum power rating of a capacitor motor is 1.7 kW, because already at a power rating of 1 kW the cost of the capacitor becomes comparable with that of the motor itself.

In comparison with three-phase machines, single-phase motors are less economical, have a lower efficiency and a lower power factor.

10-13. Losses in and Efficiency of an Induction Motor

The losses in an induction motor, as can be seen from the power diagram of Fig. 10-25, are made up of the copper loss in the stator winding, $P_{c,1}$; the stator iron loss, P_i; the copper loss in the rotor winding, $P_{c,2}$; and friction and windage (mechanical) loss P_{mech}.

Then the primary (input) power is

$$P_1 = \sqrt{3}\, V_1 I_1 \cos\varphi \qquad (10\text{-}13)$$

and the output power (on the motor shaft) is

$$P_2 = P_1 - (P_{c,1} + P_i + P_{c,2} + P_{mech}) \qquad (10\text{-}14)$$

Hence the efficiency of the motor is

$$\eta = (P_2/P_1) \times 100\% = [(P_1 - \Sigma P)/P_1] \times 100\% \qquad (10\text{-}15)$$

Motors have a maximum efficiency at rated or nearly rated load. Figure 10-28 shows the load characteristics of an induction motor, plotted in relative units (per-unit system). As is seen, they are not unlike the characteristics of a shunt-wound d.c. motor.

10-14. Synchronous Machines

At present-day power stations, mechanical energy is converted to electricity most exclusively by synchronous generators.

In a synchronous generator, the stator is similar to that of an induction machine (see Figs. 10-1, 10-7 and 10-29), and the rotor driven by a steam or water turbine at a constant speed carries a field winding energized with a direct current, I_f, as in d.c. machines.

The exciting magnetic field Φ_{exc} established by this current is revolving at a constant speed, n, and induces in the three-phase rotor winding an emf which is given by the already known equation

$$E_0 = 4.44 f w \Phi_{exc} k_0$$

If we connect the stator winding across a resistance, currents I_A, I_B and I_C will be induced in the phases of the winding, and the mmfs established by these currents, namely F_A, F_B and F_C, will combine to produce the resultant mmf, F, as has been shown in Sec. 10-2. This mmf establishes a flux in the stator or armature, Φ_a, which revolves at the same speed as the rotor. Since the field speed is the synchronous speed, and both the rotor and field revolve at this speed, the machines utilizing this feature are called *synchronous*.

In a synchronous machine, the speed n, the frequency f of the stator current and the number of pole pairs p are uniquely connected by a relation of the form

$$n = 60f/p$$

At $f = 50$ Hz and $p = 1$, 2 or 3, n_1 will be 3000, 1500 or 1000 rpm, respectively. Synchronous generators driven by steam turbines at $n = 3000$ to 1500 rpm are called *turbogenerators*. Generators driven by hydraulic turbines are called *hydrogenerators*. At power stations built on rivers flowing in plains or utilizing the runoff from water reservoirs, hydraulic turbines are in the low-speed class, and a large number of poles have to be used in order to generate current at commercial frequency ($f = 50$ Hz in the Soviet Union). For example, the 225-MW, 15.75-kV hydrogenerators operating at $\cos \varphi = 0.85$ at the Bratsk Hydro have

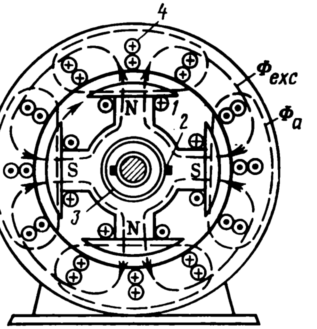


Fig. 10-29. Sketch of a synchronous generator

$p = 24$ and are designed for $n = 125$ rpm. At the Krasnoyarsk Hydro, the hydrogenerators operate at $n = 93.8$ rpm and have $p = 32$.

Figure 10-29 shows the stator and rotor of a salient-pole generator. The stator slots, *4*, hold the sides of a two-layer winding similar to that already discussed (see Fig. 10-7). A four-pole rotor carrying the field winding, *1*, is driven by a prime mover (not shown in the diagram). The exciting current is conveyed to the field winding via slip-rings, *3*, and brushes, *2*, from a separate d.c. machine called an *exciter*, or from a controlled rectifier.

An external appearance of the nonsalient-pole rotor of a turbogenerator is shown in Fig. 10-30, and a cross-sectional view of the rotor, less its winding, in Fig. 10-31.

The no-load characteristic of a generator $E_0 = f(I_f)$ at $f =$ constant and $I = 0$, and also the load (external) characteristic at $I_f =$ constant and $\cos\varphi =$ constant have the same shape as for a separately-excited d.c. generator (see Fig. 4-22). However, the percent voltage regulation

$$\Delta V = \frac{E_0 - V_n}{V_n} \times 100\%$$

of a synchronous generator is as high as 20 to 40% of the rated voltage V_n.

This is because in a synchronous machine, as in a d.c. machine, the armature reaction flux Φ_a, shown closed across a pole (see Fig. 10-29), has some of its path completed along the poles, against the exciting flux Φ_{exc}, when the

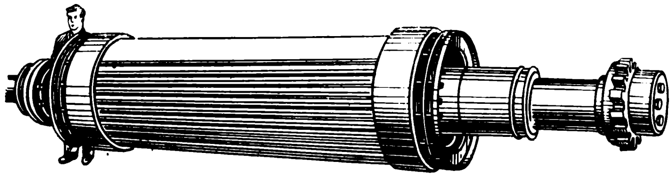


Fig. 10-30. External appearance of a turbogenerator rotor

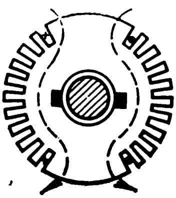


Fig. 10-31. Rotor of a turbogenerator, less its winding, and a coil of the field winding

machine is operating at a lagging current, that is, at cos $\varphi <$ < 1. As a consequence, the resultant flux, $\Phi = \Phi_{exc} - \Phi_a$, is strongly reduced, and so are E_0 and V

A synchronous machine can operate as a motor to drive mechanisms running at a constant speed, such as pumps and air blowers. In these applications, they can successfully replace induction motors at power ratings as high as hundreds or even thousands of kilowatts. Having an overload capacity, T_m/T_n, ranging from 1.8 to 2.5, a synchronous motor offers an added advantage in that it can operate at a power factor (cos φ) of unity. Figure 10-32 shows an arrangement used to start a synchronous motor. In addition to the field winding, *1*, the pole-pieces carry a short-circuited (shading) coil, *4*, similar to that of an induction motor. Prior to starting, the field winding is connected across a resistor, *3*, by a switch, *2*. The stator, *5*, is connected by a knife-blade switch, *6*, to the supply line, and the revolving magnetic flux of the stator induces currents in the shading coil and brings up the rotor, as it does in an induction motor, to a speed n_2 nearly equal to n_1. For the rotor to revolve at n_1, that is, at synchronous speed, a constant current must be maintained in the field winding. To this end, the switch is thrown down in contact with the terminals of an exciter, *7*, and the rotor automatically pulls into synchronism, after which load may be put on the motor

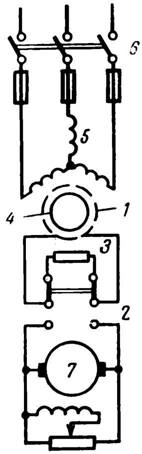


Fig. 10-32. Starting arrangement for a synchronous motor

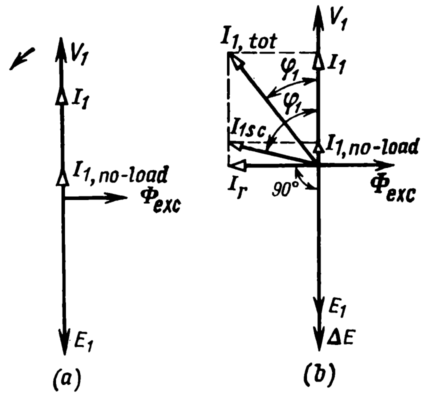


Fig. 10-33. Vector diagrams explaining operation of an overexcited synchronous motor

The vector diagram of such a motor is shown in Fig. 10-33. The revolving flux, Φ_{exc}, of the rotor induces a counter-emf, E_1 in the stator winding. If we neglect the resistance of the winding, assuming that $I_1 z_1$ approximates $I_1 x_1$, then the supply voltage is $\overline{V}_1 = -(\overline{E} + \overline{I}_1 x_1) =$ (approx.) $-\overline{E}_1$ (Fig. 10-33*a*). At no-load, $P_{no\text{-}load} = \sqrt{3}\, V_1 I_{1,no\text{-}load} \times \times \cos \varphi_{no\text{-}load}$ is very small and equals the no-load loss in the motor. The no-load current, $I_{1\ no\text{-}load}$, is an active current and is also small in magnitude, while $\cos \varphi_{no\text{-}load}$ may be equal to unity at sufficient excitation. As load on the shaft is increased, the current rises to I_1, while remaining an active one.

If we raise the exciting current I_{exc} (Fig. 10-33*b*), the exciting flux, Φ_{exc}, will rise, and the emf will go up to $E_1 + \Delta E$. Then an additional current begins to flow in the stator winding, given by

$$I_r = \Delta E / x_1$$

This is a purely reactive current, because the impedance of the stator winding is $z_1 = \sqrt{r_1^2 + x_1^2} =$ (approx.) x_1.

The current I_r is in quadrature lagging with ΔE (Fig. 10-33*b*) and in quadrature leading with V_1, so the total current of the motor $I_{1,tot}$ leads V_1 in phase by an angle φ_1.

Quite often, a synchronous motor is used as a synchronous condenser, when it does not pull any mechanical load and has a leading current, $I_{1,s.c.}$ (Fig. 10-33*b*). When a synchronous condenser is connected in a network with an inductive load, it corrects the power factor, that is, creates conditions very near those existing at series (current) resonance (see Sec. 6-10).

In comparison with static capacitors which are also used for power factor improvement, synchronous condensers permit adjustment of the leading current, but their power loss is greater than in static capacitors and they need attendance.

At power ratings of a few hundred watts, synchronous motors have no field winding. In such a modification, they are called *repulsion* motors and are used to drive mechanisms that must operate at constant speed (in sound motion-picture equipment, remote control, etc.).

10-15. The D.C./A.C. Commutator Motor

Figure 10-34 illustrates the manner in which d.c./a.c. commutator motor is connected in circuit. If the "+" and "—" terminals of such a motor are connected to a d.c. supply line, it will operate as a conventional series-wound motor discussed in Sec. 4-17. As will be recalled, if we change the direction of current flow in the armature and

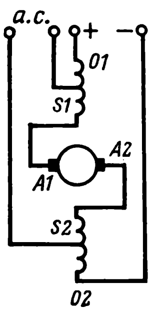


Fig. 10-34. Connection and circuit of an a.c./d.c. motor

field winding at the same time, the direction of torque will remain unchanged. As a consequence, the motor will be able to operate normally, if connected to an a.c. supply line.

However, with an a.c. supply, if no measures are taken, a good proportion of power would be dissipated as heat in the magnetic circuit, if it were made solid, and the field winding when energized with a.c. would have a high inductive reactance. To avoid this, the frame and poles of a universal machine are assembled from electrical-sheet steel punchings, like the armature core. They have no commutating poles. To bring down the inductive reactance only part of the field winding is energized in a.c. operation (Fig. 10-34).

The performance of a commutator motor in operation on a.c. is somewhat inferior to that on d.c.

Chapter Eleven Electric Drive and Control Equipment

11-1. Electric Drive

The term *electric drive* applies to an assemblage consisting of an electric motor, control equipment, and a transmission from the motor to the driven machine. In conjunction with automatic control, electric drive raises the efficiency of the power plant, improves the quality of the product, and makes labour less arduous.

A crucial matter as regards drives is the correct choice of the power rating for the motor that drives the tool (here, the word "tool" is used in a wider sense and may apply to a machine-tool, a pump, a blower, a crane, etc.). If underrated, the motor will overheat due to overloading, and its insulation may break down. If overrated, the motor will make the complete plant too expensive and reduce its efficiency and power factor. As already noted, the overload capacity of a motor is expressed as the ratio T_m/T_n; it designates the physical limit of power that a motor can develop for a short time interval. To avoid sudden stoppage, the load torque on the shaft of a motor, that is, the torque due to the driven machine, ought not to exceed T_m.

On the average, T_m/T_n is in the range from 1.8 to 2.5, being 2.3 to 3.3 for crane-driving induction motors and 1.8 to 2.5 for synchronous motors. The limit of overload on d.c. motors is set by sparking at the commutator, because overload impairs commutation to a point where a ring fire may develop around the commutator. In view of this factor, their overload capacity is set at T_m/T_n = (approx.) 2.5, and for crane motor at 3.0 to 4.0.

11-2. Temperature Rise of Electrical Machines

Electrical machines may operate in continuous, short-time and intermittent duty (see below). The values of maximum torque only apply to short-time duty. The rated torque at which a machine can operate in continuous duty depends on the limiting temperature specified for its winding insulation. Because the heat build-up in a winding depends on the load current squared, it is obvious that the rated power and current will be higher for a machine whose insulation has a higher limiting temperature.

If a machine is operated at a temperature exceeding that specified below, the service life of its insulation and, as a consequence, of the machine itself will be cut down. In contrast, operation at a reduced temperature will extend the service life of both.

As soon as a machine reaches a temperature exceeding ambient temperature, it begins to give up heat to the surroundings, and this heat flow will rise with increasing difference in temperature between the machine and the surroundings. The temperature of a machine ceases rising when the amount of heat generated in the machine becomes equal to that given up to the surroundings.

In international practice, the insulating materials used in electrical machines are divided according to the limiting temperature into seven classes.

1. Class Y. Fibrous materials made from cellulose, cotton or silk, not impregnated with or immersed in a liquid electric insulating material. Their limiting temperature is 90°C.
2. Class A. The same materials as in class Y, but impregnated with or immersed in a liquid electrical insulating material. Their limiting temperature is 105°C.
3. Class E. Some synthetic inorganic films. Their limiting temperature is 120°C.
4. Class B. Materials based on mica, bonded by or impregnated with organic resins or varnishes. Their limiting temperature is 130°C.
5. Class F. Inorganic materials fabricated from mica, glass fibre or asbestos in combination with synthetic binders or fillers. Their limiting temperature is 155°C.

6. Class H. The same materials as in class F, but in combination with silicone binders and fillers. Their limiting temperature is 180°C.

7. Class C. Mica, porcelain, glass, and quartz with inorganic binders or without any. The limiting temperature is 180° or so.

At an ambient temperature of +35°C, the temperature rise for the steel cores and other parts in contact with windings using class A and B insulation ought not to exceed 65 and 85 degrees C. Under the same conditions, the temperature rise for the slip-rings is 70 and 90 degrees C, and for commutators the respective figures are 65 and 85 degrees C. For journal (sleeve) bearings, the temperature ought not to exceed 80°C, and for anti-friction bearings, 95°C.

The limiting temperature, ϑ, of a motor is the sum of its temperature rise, τ (which is the excess above ambient temperature) and the ambient temperature, $\vartheta_0 = 35$°C, that is, $\vartheta = \tau + \vartheta_0$. As ϑ_0 decreases, a motor is likely to be overloaded.

11-3. Selection of Power Rating for a Motor in Continuous Duty

If a motor is pulling a constant load, its rated power must be equal to or somewhat exceed the power required to ensure the normal operation of the driven machine (which may be a machine-tool, a pump, a fan, a crane, and the like).

The power ratings of driven machines are found by suitable equations and are given in reference books.

To determine the power of a motor operating in continuous duty, one need have a load curve, $I = f(t)$ (Fig. 11-1). For design purposes, the gradually changing curve is replaced by a stepped broken line, assuming that in time t_1 the current in the motor is I_1, in time t_2 it is I_2, etc. The varying current is replaced by an *equivalent* current, I_e, that is, one which will give up as much heat as the stepwise varying current over an operating cycle, t_c. Then

$$I_e^2 r t_c = I_e^2 r\,(t_1 + t_2 + \ldots + t_n)$$
$$= I_1^2 r t_1 + I_2^2 r t_2 + \ldots + I_n^2 r t_n$$

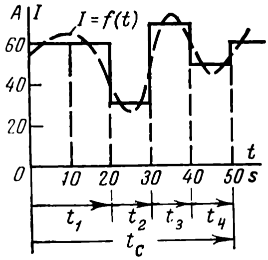


Fig. 11-1. Load curve used to select power rating for a motor operating in continuous duty

and the equivalent current is

$$I_e = \sqrt{\frac{I_1^2 t_1 + I_2^2 t_2 + \ldots + I_n^2 t_n}{t_1 + t_2 + \ldots + t_n}} \tag{11-1}$$

The rated current of a motor must be equal to or exceed the equivalent current, that is, $I_n \geqslant I_e$.

For synchronous motors and shunt-wound d.c. motors operating at a constant exciting flux,

$$T = c_T \Phi I \sim I$$

This is why motors can be selected on the basis of their equivalent torque

$$T_e = \sqrt{\frac{T_1^2 t_1 + T_2^2 t_2 + \ldots + T_n^2 t_n}{t_1 + t_2 + \ldots + t_n}} \tag{11-2}$$

If a motor has a flat characteristic and, as a consequence $P = T\omega \sim T$, it can be selected on the basis of its equivalent power

$$P_e = \sqrt{\frac{P_1^2 t_1 + P_2^2 t_2 + \ldots + P_n^2 t_n}{t_1 + t_2 + \ldots + t_n}} \tag{11-3}$$

If the load curve shows intervals of short-time overloads, the selected motor must be tested for its overload capacity.

11-4. Selection of Power Rating for a Motor Operating in Short-Time Duty

If a motor is allowed to run during an interval t_{st} during which its temperature does not reach a steady-state value, and it has time to cool to ambient temperature when turned off, it is said to operate in short-time duty (this applies

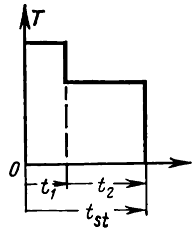


Fig. 11-2. Load curve for a motor operating in short-time duty

to the chucks of metal-cutting machines, water-lock drives, bascule bridges, etc.). Then, in accordance with its load curve (see Fig. 11-2) it is assumed that

$$t_1 + t_2 + \ldots + t_n = t_{st}$$

and T_e is found by Eq. (11-2). Then a suitable catalogue is looked up for a motor that is designed to operate in short-time duty for intervals equal to t_{st} and whose torque is equal to $T_n \geqslant T_e$. The motor thus selected is then tested for instantaneous overload current, so that I_m/I_e is allowable for the motor in question.

11-5. Selection of Power Rating for a Motor in Intermittent Duty

During operation in intermittent duty, a motor does not reach its steady-state temperature; neither does it cool to ambient temperature when stopped (Fig. 11-3). This applies to the motors used in lifts, hoists, excavators and similar applications. For such motors, data sheets state their duty factor. The duty factor is the ratio of the "on" time to the duty cycle, t_c, which is the sum of the "on" time and the "off" time (or period), t_{off}:

$$DF = \frac{t_1 + t_2 + \ldots + t_n}{t_1 + t_2 + \ldots + t_n + t_{off}} \times 100\% \qquad (11\text{-}4)$$

As the duty factor increases, the rated power decreases for a motor of the same size. As a consequence, a motor designed to operate for 25% of the cycle at rated power must not be left under load for 60% of the cycle at the same power.

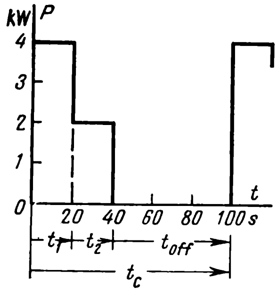


Fig. 11-3. Load curve for motor operating in intermittent duty

In the Soviet Union, electric motors are built for duty factors of 15, 25, 40 and 60%, the nominal figure being 25%. If the duty cycle does not exceed 10 min, a motor is designed for an intermittent duty, using Eq. (11-3). Then, in the case of crane motors, the rated power is found from a catalogue for the specified duty factor. If the duty factor thus found is other than standard, P_{e1} found by Eq. (11-3) must be re-computed in terms of the standard duty factor

$$P_{e2} = P_{e1} \sqrt{DF_1/DF_2} \qquad (11\text{-}5)$$

11-6. Knife-Blade Switches

Knife-blade switches are used for manually starting electric machines and closing d.c. and a.c. circuits at voltages up to 500 V and currents up to 1000 A. Knife-blade switches may be one-, two-, and three-pole.

Figure 11-4*a* and *b* shows a simple knife-blade switch and its diagram symbol (accepted in the Soviet Union). It has contact blocks, *2*, set up on an insulating base, *1*, to which are connected the wires, *3*, of the circuit being controlled. The metal blades, *4*, of the switch are attached to the lower hinge jaws. Figure 11-4*b* shows a three-phase knife switch in which the blades are joined by a cross bar, *5*, so that the blades are cut into their respective break jaws all at the same time. The switch is operated by a handle, *6*.

When a knife-blade switch opens an energized network, an arc, *7*, strikes between the upper break jaws and the blades. This arc might melt or burn the contacts. Because

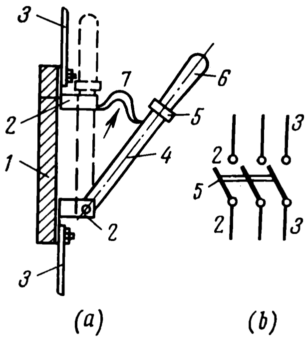


Fig. 11-4. Simple knife-blade switch

of this, the knife-blade switch shown in Fig. 11-4 can only be used as a disconnecting or isolating switch in a.c. circuits operating at not over 220 V, that is, as a switch opening a circuit at no-load.

So that a knife-blade switch can open a circuit under load, it is fitted with arcing contacts (Fig. 11-5). When

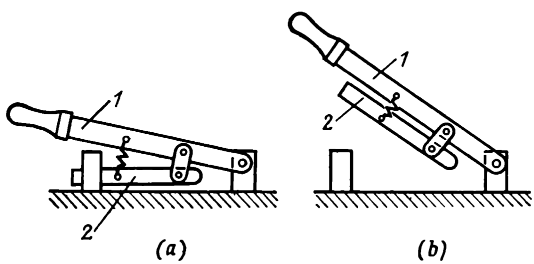


Fig. 11-5. Quick-break knife-blade switch

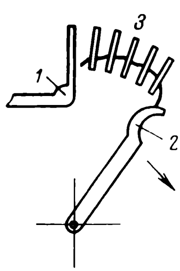


Fig. 11-6. De-ion grid

the circuit is being opened, the first to break are the main blades, *1*. They are followed by auxiliary, or arcing, blades, *2*, operated by a spring. Because the action is very quick, this type is called a quick-break switch. In this form of switch, the arc is extended and quenched.

Switches used to open a.c. circuits operating at 380 and 500 V and d.c. circuits operating at 220 V and more under load are fitted with arc-extinguishing features. One such device is a *de-ion grid* (Fig. 11-6). The de-ion grid, *3*, is assembled of steel bars placed above the main contacts, *1* and *2*, and enclosed in an arc-chute made of asbestos cement, ceramic materials and the like. The arc that strikes when the switch is being opened is driven by electrodynamic forces and by the flow of hot air into the de-ion grid where it is split up into a multiplicity of small arcs. An arc requires a definite voltage for its existence. If it is divided into a number of smaller arcs such that the voltage between contacts *1* and *2* is less than the voltage required to maintain the individual arcs, the latter will rapidly die out. At voltages of 220, 380 and 500 V, de-ion grids and switches are completely enclosed in protective cases.

11-7. Packet Switches

The size of switchgear used for manually closing and opening electric circuits operating at 220 V and 380 V d.c. and a.c. and a direct current of 10 to 400 A can be reduced through the use of packet switches (Fig. 11-7*a*).

The wires of the supply line are connected to stationary contacts, *3*, set up on fixed decks of an insulating material, *2*. When a quadrangular shaft is rotated by a handle, *5*, the movable contacts, *4*, arranged on the shaft can take up two positions displaced from each other by 90°, as shown in Fig. 11-7*b* and *c*. These are double (or split) contacts so arranged that they grip the stationary contacts from above and below. In the position shown in Fig. 11-7*b*, the stationary contacts are connected together ("shorted") by the movable contacts. In the position shown in Fig. 11-7*c*, they are disconnected from each other ("open-circuited"). This reasoning applies to the three decks, *2*, stacked one upon another.

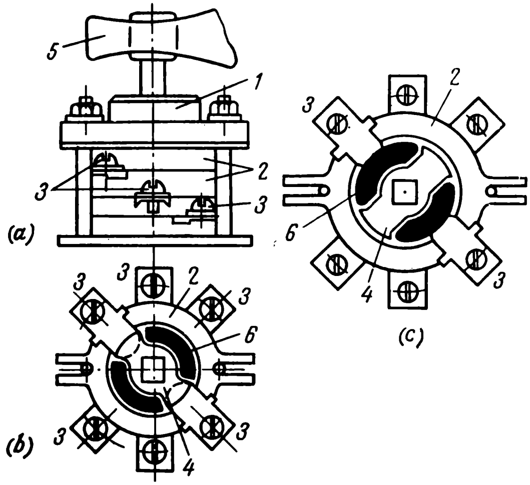


Fig. 11-7. Packet switch

There are also hard-fibre disks, *6*, lying in the same plane and rotating together with the movable contacts, *4*. In the position shown in Fig. 11-7*c*, the discs clamp the stationary contacts from above and below. The function of the hard-fibre discs is to extinguish the arc that strikes when the contacts are opened. When the arc strikes, it decomposes the hard fibre with the liberation of hydrogen, carbon dioxide and water, which fact leads to a fast extinction of the arc. As the handle, *5*, is rotated, a spring mechanism installed under the cover, *1*, forces the contacts apart, thereby cutting down the time required for opening the circuit.

11-8. Starting and Control Rheostats for Electric Motors

The functions that rheostats perform and the manner in which they can be connected in a circuit are discussed in Chapters 4 and 10. Basically, rheostats are used to control current and are assembled from standard elements which are

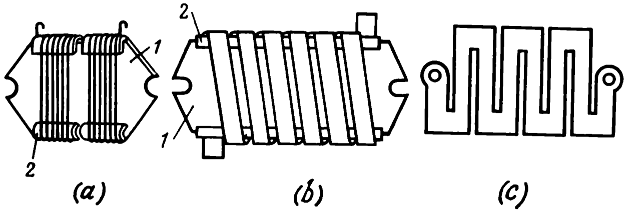


Fig. 11-8. Resistance elements

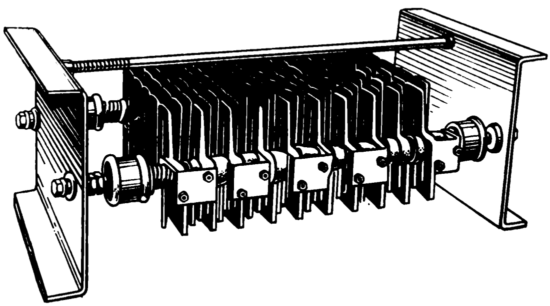


Fig. 11-9. Resistor assembled from cast-iron sections

resistors set up on an insulating baseplate. The resistors are wound with constantan, manganin, nichrome, fechral, and steel wire or ribbon, or they may be cast-iron sections. Some of the element types fabricated from wire and ribbon are shown in Fig. 11-8*a* and *b*. They are wound on and insulated from a steel former, *1*, by porcelain or steatite riders, *2*. Figure 11-8*c* shows a cast-iron resistance element such as are used in resistance boxes (Fig. 11-9).

In contrast to resistance boxes, rheostats have a switching device to control their resistance (Fig. 11-10).

When the handle, *2*, is turned on the cover, *1*, a shaft, *3*, is rotated together with its brushes, *4*. Leads from the sections, *6*, are taken to terminals arranged in a circle on an insulating baseplate, *5*. The rheostat terminals are located under a box, *7*, and the entire structure is immersed in an oil-filled tank for cooling.

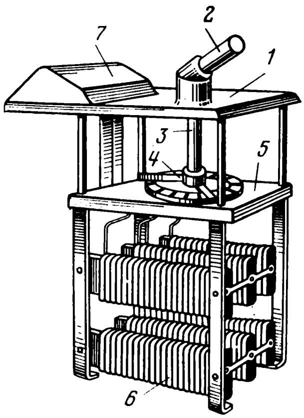


Fig. 11-10. Rheostat

Apart from the most commonly used metal rheostats, there are also liquid rheostats such as, for example, used to start slip-ring induction motors. A liquid rheostat is a tank holding an amount of electrolyte (usually, water with the addition of 8 to 10% of soda), and metal plates insulated from one another and dipped in the tank. This form of rheostat provides continuous control. The resistance of a liquid rheostat is proportional to the separation between the plates and inversely proportional to the immersed surface of the plates.

11.9. Control Switches

Control switches* are devices which are used to bring resistors in and out of circuit in order to start, stop or reverse motors. A control switch (the switching mechanism proper) is usually built into a case which is installed separately from the box-enclosed resistors it switches (see Fig. 11-9). In contrast to a multipole knife-blade switch, a control switch closes and opens different circuits at different times, in a predetermined sequence. Figure 11-11*a* shows a top view of one element of a *cam-type control switch*.

A vertical frame, *3*, mounts a plastic insulator, *2*, which gives support to a stationary switch contact, *1*. A wire,

*Sometimes, they are also called controllers, master switches or switch starters.— *Translator's note.*

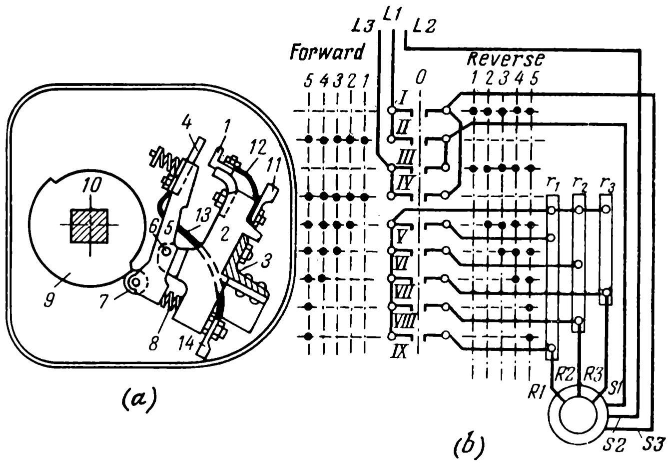


Fig. 11-11. Cam-type control (or master) switch

12, conveys to this contact the current from a cable lug, *11*, in the controlled circuit. The base of a movable contact, *4*, is another plastic insulator, *5*, which is free to rotate on a pivot, *6*. This contact is connected to the second cable lug, *14*, of the controlled circuit by a wire, *13*. The movable contact, *4*, is loaded by a spring, *8*, which tends to force it against the stationary contact, *1*. The movable and stationary contacts are closed and opened, as the shaft is rotated, by a cam, *9*, mounted on a vertical shaft, *10*, opposite a roller, *7*. A typical control switch has a set of cam segments mounted along the shaft, *10*, and their number depends on the manner in which the associated motor is started.

In the starting scheme for a slip-ring induction motor shown in Fig. 11-11*b*, the control switch (or the switch starter) has nine switch elements (*I*, *II*, *III*, *IV*, *V*, *VI*, *VII*, *VIII* and *IX*). Schematically, they are shown in the same figure. The cover of the control switch case bears numerals which designate the positions (*1*, *2*, *3*, *4*, *5*, forward, reverse) into which the handle rotating the shaft, *10*, can be placed. In the diagram of Fig. 11-11*b*, the closed contact elements are marked by solid dots in the respective positions.

In the "0" position, the stator is connected to the supply line at terminal *S2*, the rotor circuit contains resistors r_1, r_2 and r_3 connected in a star, and the motor is de-energized. In the "*1* — forward" position, switch elements *II* and *IV* close, the stator is energized, and the motor starts rotating at full resistance in the rotor circuit. As the handle is moved to the "*2*" position, the next, switch element, *V*, closes and shunts part of the resistor r_1. As the handle is consecutively moved to the "*3*", "*4*" and "*5*" position, switch elements *VI*, *VII*, *VIII* and *IX* close in turn, thereby shunting the resistors in all the phases of the rotor.

When the handle is turned in the reverse direction, switch elements *I* and *III* operate instead of elements *II* and *IV*, so that two phases in the stator circuit are interchanged, and the motor is reversed. The rotor circuit is switched in the same sequence.

11-10. Fuses

Fuses are devices which protect circuits and apparatus against short-circuit currents. In the case of a short-circuit, the fuse link made of copper, zinc, lead or silver melts and opens the circuit.

Zinc, lead and its alloys melt at a low temperature (200° to 240°C), but they have a low conductivity, so fuse links made from them must of necessity have a large cross-sectional area. Copper fuse links have a good conductivity, but they melt at a fairly high temperature (960° to 1080°C). Silver fuse links are expensive and are used at voltages over 1000 V and low currents. Most commonly, use is made of copper fuse links silvered for protection against oxidation, with lead or tin globules called "dilutants" soldered onto the silver plating (see below).

Figure 11-12*a* and *b* shows an external appearance and a sectional view of a plug-type fuse. A plug fuse consists of a porcelain container enclosed in a thin-walled threaded metal sleeve, *4*, called the base, which screws into a threaded fuse receptacle, *3*. From the terminal, *1*, the current path is through a metal plate, *2*, fuse receptacle, *3*, threaded (or screw) base, *4*, fuse link, *5*, contact, *6*, to the opposite terminal, *7*. When the fuse (Fig. 11-12*c*) should be replaced,

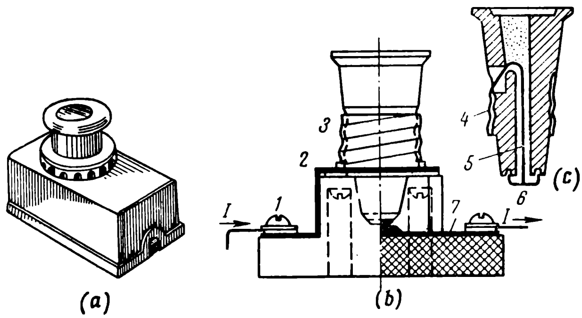


Fig. 11-12. Plug fuse

it can be unscrewed by the container, without having to touch its current-carrying parts. Plug fuses are manufactured for voltages up to 380 V and currents up to 60 A.

Figure 11-13*a* shows a cartridge fuse. It consists of a hard-fibre tube, *3*, enclosing a fuse link, *6*, and installed like the blade of a knife switch in contact jaws, *2*, set up on an insulating baseplate, *1* (Fig. 11-13*b*). The tube is tightly stopped at both ends by brass terminal caps, *4*, to which are made knife-blade contacts, *5*. They are used for connection of the fuse-link terminals and insertion into the jaws, *2*. A short circuit causes the link to melt at several narrow places at a time, which fact speeds up the extinction of the arc. Also, the arc decomposes some of the hard-fibre into a mixture of hydrogen, carbon dioxide and steam. The pressure inside the tube rises to 10^7 Pa (100 kgf/cm^2), and the arc is rapidly quenched.

Figure 11-14 shows a sand fuse used for voltages up to 500 V at currents from 100 to 600 A. Basically, it consists of a fuse case, *5*, completely filled with quartz sand, *6*, which surrounds fuse links, *7*, made of fine copper ribbons and attached to knife-blade contacts, *3*, with which the fuse is held in contact jaws, *2*, set up on an insulating baseplate, *1*. The jaws keep the sand filler in place; the fuse case is stopped by a cap, *4*. Midway along their length, the fuse links have tin straps, *9*, soldered to them. When an overload occurs, the tin straps melt and enable the copper links to blow at a faster rate. In the case of a short-circuit, the links blow at perforations, *8*.

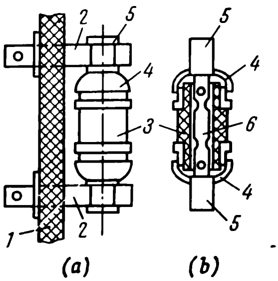


Fig. 11-13. Cartridge fuse

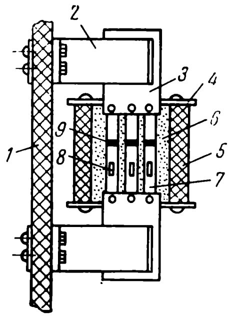


Fig. 11-14. Sand fuse

In circuits carrying currents from 15 to 60 A, use is made of sand fuses having an integral case. When a fault occurs, an arc is formed; it comes in contact with the fine sand (grains) which rapidly cools it, so the arc is de-ionized and goes out quickly.

11-11. Automatic Air Circuit Breakers

An automatic air circuit breaker is a switching apparatus used for manually closing and automatically or remotely opening the associated power circuit or network under load. In a.c. networks, they are installed if the voltage is up to 500 V, while in d.c. networks they are used even at higher voltages, if switching operations occur not more than a few times in 24 hours. Automatic circuit breakers are fitted with a release (also known as a trip) mechanism or simply a release.

A release (Fig. 11-15*a* through *c*) has a moving system actuated by devices utilizing electromagnetic or thermal effects of current.

Figure 11-15*a* shows an *overcurrent release*. When the load current exceeds a predetermined value, the electromagnet, *1*, attracts the armature, *2*, and releases a latch, *3*, while spring, *4*, forces the main contacts, *5*, of the circuit breaker to separate. Figure 11-15*b* shows an *undervoltage release*. Here, the electromagnet, *1*, releases its armature, *2*, in the case of a reduction in voltage or a complete voltage failure,

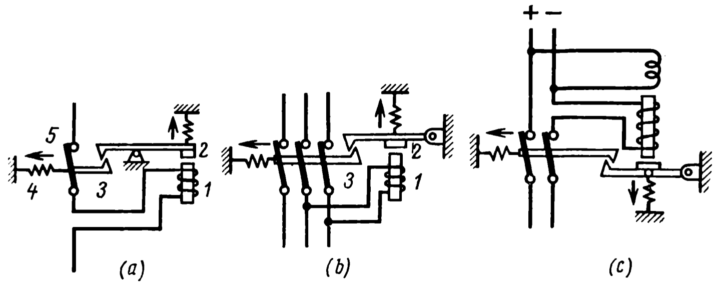


Fig. 11-15. Circuits of releases

so that the latch, *3*, is released, too, and the power circuit is interrupted. In a *reverse-current release* (Fig. 11-15*c*) the latch is released when the direction of current flow is reversed and the total mmf of the series (current) and shunt (voltage) coils changes sign, so the power circuit is opened. The time required for a circuit-breaker to open t_o is 0.05 to 0.025 s. An undervoltage circuit-breaker is usually set to operate (to open) at $V = 0.4\ V_n$.

11-12. Contactors

A contactor is a switch capable of closing and opening the associated circuit(s) up to 1500 times an hour. They are used in d.c. and a.c. circuits operating at up to 1000 V. They give no short-circuit or overload protection, so they must be used in conjunction with protective devices. Contactors are set to operate at 0.85 to 1.03 the rated voltage V_n and will automatically de-energize the associated electric installation, should the voltage drop to 0.5 or 0.6 of its rated value.

Figure 11-16 shows an a.c. three-pole contactor for currents from 20 to 600 A. Its operate time ranges from 0.05 to 0.1 s.

When the "start" button in the control circuit is pressed, the coil of a solenoid, *8*, is energized, the core, *7*, attracts the armature, *3*, and an insulated shaft, *1*, rotates in bearings (not shown in the diagram). This causes the movable and stationary contacts (*2* and *6*, respectively) in the power

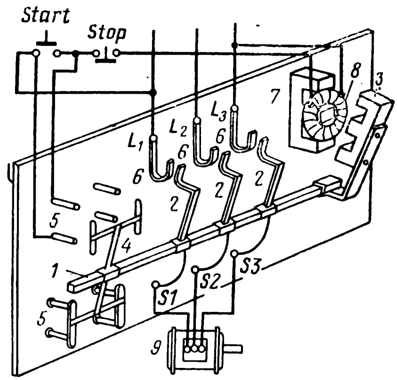


Fig. 11-16. Arrangement of a contactor

circuit to close, and the motor, *9*, is set in motion. At the same time, a bridge, *4*, rotates to separate the lower interlock contacts, *5*, and to close the upper interlock contacts—these interlock contacts lock the contactor in the energized position, and now the "start" button may be released. According to their action, the upper interlock contacts are called *make* (or *normally-open*, N.O.) contacts, and the lower interlock contacts are called *break* (or *normally-closed*, N.C.) contacts. When the "stop" button is pressed, the solenoid, *8*, is de-energized, and the main contacts, *2* and *6*, break. The main contacts are enclosed in arc-chutes. In a.c. contactors, the magnetic flux in the solenoid, *8*, periodically passes through zero value, and this might cause chatter and hum. To avoid this, a shading ring (which is a shorted turn) is placed around a portion of the coil core, in which a current is induced as in the secondary winding of a transformer, this current being shifted in phase relative to the current in the solenoid, *8*. At the instants when the main flux passes through zero, the flux due to the shading ring prevents the armature to drop out.

11-13. Relays

The forms of electric devices we have so far discussed are designed for operation in power circuits and can switch currents over 5 to 10 A. Devices designed for lower currents and for operation in control (rather than power) circuits

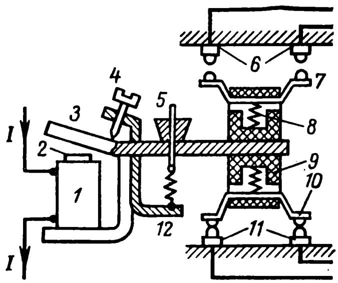


Fig. 11-17. Arrangement of an electromagnetic current relay

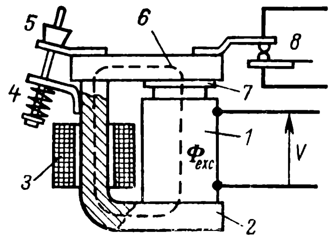


Fig. 11-18. Arrangement of a time relay

at low voltage are called *relays*. They have simple control contacts and no arc-extinguishing features.

More specifically, a relay is an element in a d.c. or a.c. control circuit, which operates instantaneously or with some delay after the controlled quantity (which may be current, voltage, and so on) reaches a preset value. The minimum current causing a relay to pick up or to drop out (that is, its contact to make or break) is called its *operate current*. The maximum current causing a relay to reset (that is, its contact to take up their initial condition) is called its *release* (or *drop-out*) *current*.

Figure 11-17 illustrates the operation of an *electromagnetic current relay*. When the current in the coil, *1*, placed in the power circuit controlled reaches the operate value, the electromagnet armature, *3*, is attracted to a pole-piece, *2*. As a result, contacts, *6* and *7*, make, and contacts, *10* and *11*, break. The movable contacts, *7* and *10*, are mounted on a lever, *3*, by means of plastic blocks, *9*, and springs, *8*. The operate current is set by adjusting the tension of the opposing (or control) spring, *12*, with a nut, *5*, and also by varying the air gap in the solenoid with a screw, *4*.

Figure 11-18 gives a schematic representation of an electromagnetically delayed *time* (or *timing*) *relay* (sometimes called a timer). This relay releases to its de-energized position with a fairly large delay (up to 5 s) after removal of voltage from the terminals of its coil, *1*. The relay is shown in the energized position, prior to the opening of the contacts, *8*, in the auxiliary circuit.

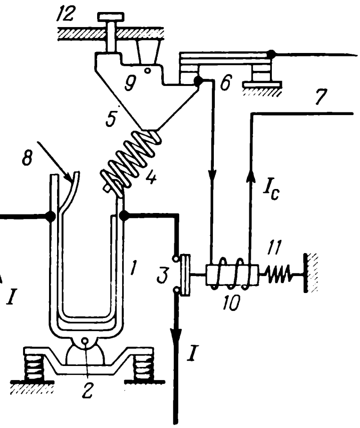


Fig. 11-19. Circuit of a thermal relay

In more detail, the time relay operates as follows. The magnetic circuit, *2*, of the relay has a shading sleeve, *3*, in which a decrease in the excitation flux Φ_{exc} establishes (by Lenz's law) a current and a flux Φ_{op}, which tend to maintain the collapsing flux Φ_{exc}. So, when the voltage V is removed, the armature, *6*, is not released by the core of the coil at once. The greater the tension of the spring, *4*, adjustable with the nut, *5*, the greater the flux Φ_{exc} at which the relay will operate and, as a consequence, the shorter its release time. This time can also be adjusted by varying the thickness of the brass plate, *7*, known as an armature stop.

There are also a large variety of *thermal relays*. In most cases, they use a bimetallic thermal element, that is, a combination of two metal strips differing in thermal coefficient of linear expansion and welded together along their length. When the bimetallic element is heated, the difference in expansion causes it to flex. In doing so, the bimetallic element actuates the device that closes or opens the associated electric circuit. The bimetallic element can be heated either directly by the current flowing through the whole or part of the strip, or indirectly, by a separate heater. A simplified presentation of such a relay is shown in Fig. 11-19.

A U-shaped bimetallic strip, *1*, is mounted on a pivot, *2* and connected in series with the main contacts, *3*, in the

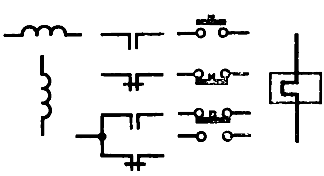
controlled power circuit. The right-hand end of the bimetallic strip bears upon a coil spring, *4*, attached to an insulating block, *5*, which carries the movable silvered auxiliary (break) contacts, *6*, placed in the control circuit. The left-hand end of the bimetallic spring is attached to the setting mechanism, *8*, which is intended to adjust the operate time of the relay by pre-flexing the strip.

When the current *I* in the power circuit reaches the operate value, the bimetallic strip flexes, and the coil spring rapidly rotates the plastic block on its pivot, *9*, thereby separating the auxiliary (N.C. or break) contacts in the control circuit. As a result, the solenoid, *10*, is de-energized, and the main contacts in the power circuit are caused by another spring, *11*, to separate. The relay is reset manually, using a button, *12*.

11-14. Motor Control Circuits

For all its clarity, the diagram illustrating the operation of a contactor in Fig. 11-16 is elaborate and difficult to show in a drawing. This is why in practice use is made of diagrams in which all elements of an apparatus are shown by symbols. Some of these symbols appear in Fig. 11-20. The switching devices are shown in the de-energized position, that is, in the absence of current in any part of the circuit or of any external forces acting on the movable parts of a mechanism.

Most often, control circuits are shown (in Soviet practice) functionally distributed, that is, in the various parts of the



(a) (b) (c) (d)

Fig. 11-20. Some circuit symbols

(*a*) relay, contactor, or magnetic starter coil; (*b*) relay, contactor, starter contacts, auxiliary normally open (N.O. or make) contacts, auxiliary normally closed (N.C. or break) contacts, auxiliary make-break contacts; (*c*) momentary buttons: normally open, normally closed, break-make; (*d*) thermal-relay heater element

circuit according to the electrical connection existing between them, rather than as actually arranged in an apparatus. The main (power) circuits are shown separately from the auxiliary (control) circuits.

In a diagram, each device is assigned a distinct letter symbol (for example, in Russian, K stands for a contactor, TT for a current transformer, etc.), the basic rule being to assign the same letter to the component elements of a device as it has in the main (power) circuit.

11-15. Control of a Three-Phase Squirrel-Cage Induction Motor by a Magnetic Starter

A magnetic contactor (that is, one operated by an electromagnet) may be used to control three-phase induction motors. In such a case, it is often called a *magnetic starter*. Its detailed diagram, based on Fig. 11-16, is shown in Fig. 11-21. In the diagram, the main (power) and auxiliary (control) circuits are shown separately. The power circuit contains fuses to protect the motor against short-circuit currents, and also the main contactor contacts and the heaters, *T1* and *T2*, of the thermal relays protecting the motor against overload.

The control circuit is a series combination of the following elements: a "start" button, Start, a "stop" button, Stop, the contactor electromagnet coil and the thermal-relay normally closed (N.C.) contacts. The "start" button is shunted by normally open auxiliary contacts which close at the same time with the main contactor contacts.

When the "start" button is pressed, it completes a closed loop in the control circuit, running from the terminal marked *L3*, via the contactor electromagnet coil, the "start" button, the "stop" button, the contactor electromagnet, the N.C. contacts of the thermal relays *T1* and *T2*, to terminal *L2*. Now the "start" button may be released, because it is shunted by the N.O. auxiliary contacts, which close the same instant as the main contacts do so in the power circuit. To stop the motor, it is enough to press the "stop" button. This will interrupt the control circuit, and the main contacts of the contactor will break.

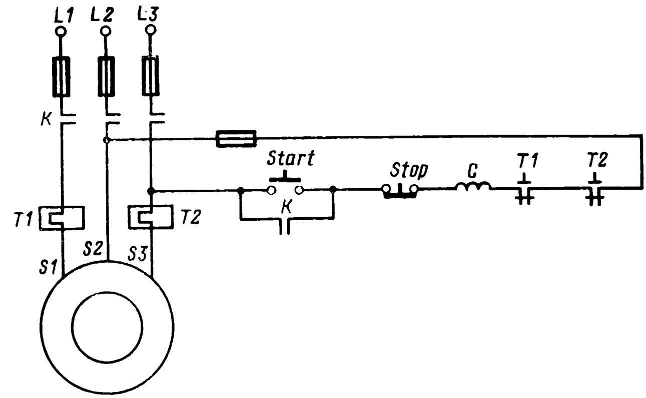


Fig. 11-21. Control of a three-phase squirrel-cage induction motor

When an overload occurs, the thermal-relay heaters cause the respective N.C. contacts to break in the control circuit, while the main contactor contacts open the power circuit. The "⊥" symbols above the relay contacts represent the manual-reset buttons to the energized position.

11-16. Control of a Three-Phase Squirrel-Cage Induction Motor by a Reversible Magnetic Starter

So that a motor can be made to run in the forward and reverse direction at will, one can install two contactors in the power circuit (Fig. 11-22), one to start it in the forward direction, "*FC*", and the other in the reverse direction, "*RC*". As in the circuit of Fig. 11-21, short-circuit protection is provided by fuses, and overload protection by thermal relays, *T1* and *T2*.

In addition to a "stop" button, the control circuit contains two more buttons to start the motor in the forward direction and in the reverse direction. They have auxiliary normally open (N.O.) contacts. There are also N.C. thermal-relay contacts and reset buttons shown above these contacts.

The motor is started forward by pressing the "*F*" button, which causes its upper contacts to break and the lower contacts to make, thereby establishing a current path in the

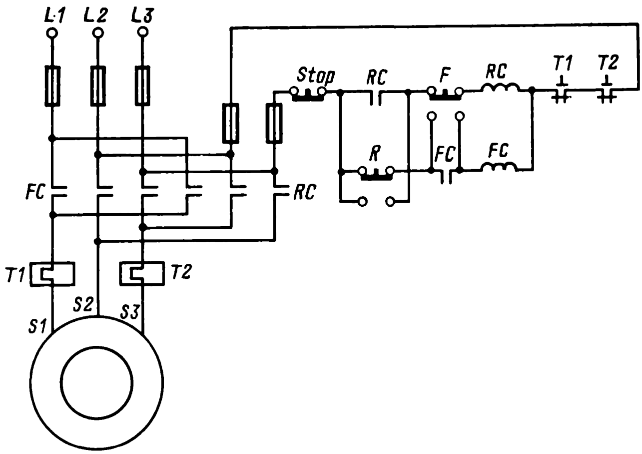


Fig. 11-22. Control of a three-phase squirrel-cage induction motor, using a reversible magnetic starter

control circuit from terminal *L3*, via the "stop" button, the closed upper contacts of the "*R*" button, the closed lower contacts of the "*F*" button, the main forward contactor coil, the N.C. thermal-relay contacts to terminal *L2*. As a result, the main forward contacts close, and the motor starts rotating. The reverse contactor, *RC*, cannot operate, because there is no current path through its coil. The same instant as the main forward contacts close, the auxiliary contacts of the "*F*" button make, and the button may be released. When the "stop" button is pressed, the supply circuit to the forward contactor coil is interrupted, and the motor comes to a stop.

If, without pressing the "stop" button, we press the "*R*" (reverse) button, its upper N.C. contacts will break before the lower contacts make. As a result, the "forward" control current path is interrupted and the "reverse" control current path is completed. As is seen, these two control paths are interlocked. In the reverse current path, the current flow is from terminal *L3*, via the "stop" button, the closed lower contacts of the "*R*" button, the closed contacts of the "*F*" button, the "reverse" contactor coil, and the thermal relays' contacts to terminal *L2*. The "reverse" contactor's main

contacts make, and the motor starts running in the "reverse" direction. Now the "*R*" button may be released, because its N.O. auxiliary contacts make at the same time with the "reverse" contactor's main contacts and lock the contactor in the energized position.

11-17. Starting of a Two-Speed Squirrel-Cage Induction Motor

A two-speed induction motor may have either two windings, each corresponding to a particular rpm, or one winding which can be switched to any of two speeds. The diagram of Fig. 11-23 shows the latter case. For the lower speed, the delta-connected stator winding is connected to the supply line at its terminals marked *S4*, *S5* and *S6*. For the higher speed, it is connected into two stars in parallel which are then connected to the supply line at the terminals marked *S1*, *S2* and *S3*. At either speed, the motor delivers the same power. This type of motor is used in the drives of metal-cutting machine-tools.

From Fig. 11-23 it is seen that the power circuit of the stator contains a low-speed contactor *LC* and a high-speed contactor *HC*, and also fuses, and the heaters of thermal relays. The control circuit contains the normally-closed "stop" button, the "*H*" (high-speed) start button, the "*L*" (low-speed) start button, the normally-open auxiliary contacts of these two buttons, the two contactors' electromagnet coils, and the normally closed contacts of the thermal relays.

When the "*L*" button is pressed, its upper contacts break and its lower contacts make. As a result, the following current path is set up in the control circuit: from the "stop" button, via the upper contacts of the "*H*" button, the lower contacts of the "*L*" button, the electromagnet coil and the normally closed contacts of the thermal relays. The main contacts make and the motor is allowed to run at the lower speed. Now the "*L*" button may be released, because the established circuit will be maintained by closure of the normally open auxiliary contacts.

When the "*H*" button is pressed, the control circuit is interrupted, the main contacts break, and so do the N.O. auxiliary contacts. Then the lower contacts of the "*H*"

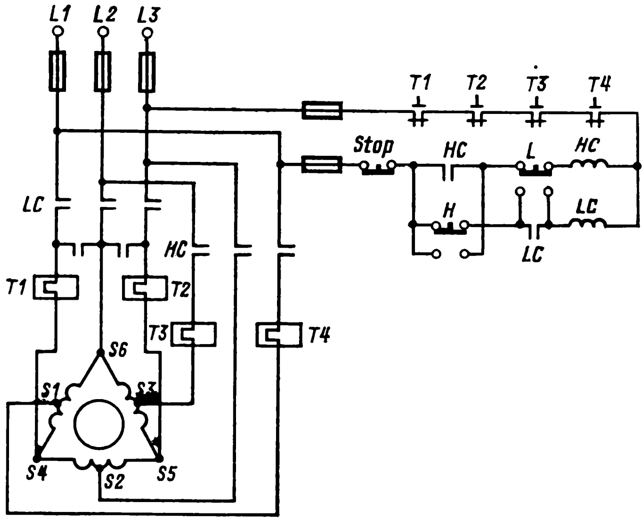


Fig. 11-23. Starting of a two-speed squirrel-cage induction motor

button make, thereby establishing the following current path in the control circuit: from the "stop" button, via the contacts of the "*H*" button, the contacts of the "*L*" button, the contactor's coil and the N.C. contacts of the thermal relays. As a result, the main contacts of the high-speed contactor close, and so do the normally opened auxiliary contacts of the "*H*" button, with the result that the motor changes to the higher speed. Now the "*H*" button may be released.

11-18. Automatic Starting of a Three-Phase Slip-Ring Induction Motor

The starting arrangement for a three-phase slip-ring induction motor is shown in Fig. 11-24. As is seen, the stator winding is connected to the supply line via a line contactor, *LC*, and the rotor is connected to three identical pairs of resistors, r_1 and r_2, connected in a star. Starting is done by means of acceleration contactors, *AC1* and *AC2*,

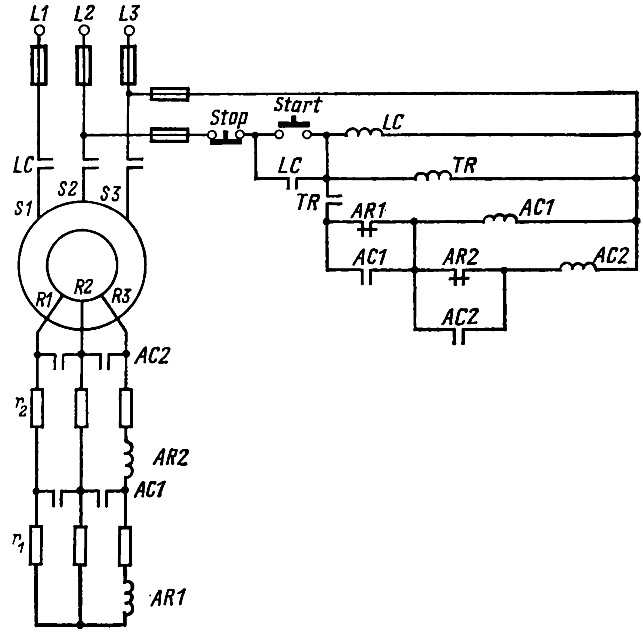


Fig. 11-24. Automatic starting of a three-phase slip-ring induction motor

acceleration current relays, *AR1* and *AR2*, and a time relay, *TR*.

When the "start" button is pressed, the line contactor coil is energized, and the line contactor's main contacts make. The line contactor's normally-open auxiliary contacts also make, thereby shunting the "start" button, and it may now be released. At the same time, the time-relay coil is also energized, and the time relay closes its contacts with a delay. By the time the time-relay contacts make, the acceleration relays, *AR1* and *AR2*, have time to operate and to open its normally closed contacts.

When, in going down, the rotor current reaches the drop-out value for the relay *AR1*, the latter's contacts make, allowing a current to flow through the coil of the acceleration contactor. As a result, the main contacts of this contactor close and short out the resistors r_1. At the same

time, the normally open contacts of the first acceleration contactor close and shunt the normally closed contacts of the first acceleration relay. The contacts of the second acceleration relay cannot make, because the reset current for that relay is lower than that for the relay *AR1*.

Some time later, after a second current inrush in the rotor, it drops to the drop-out value for the second starting step. The contacts of the second acceleration relay make, a current begins to flow in the coil of the second acceleration contactor and its main contacts close. The resistances r_2 are shorted out at the same time as the contacts of the relay *AR2* in the control circuit are shunted by the normally open auxiliary contacts of the second acceleration contactor. Now the rotor of the motor is short-circuited, and the motor is fully under way.

Chapter Twelve — Electric Power Transmission and Distribution

12-1. Industrial Distribution Networks

In most cases, industrial enterprises draw their power from a power transmission system.

According to a plant's requirements in power and some other factors, electricity is delivered to users at voltages of 110, 35, 6 or 0.4/0.23 kV*.

When a factory is a considerable distance from a power station or a substation of a power transmission system, power is transmitted at a higher voltage for economical operation. So, at the factory, the incoming voltage has first to be applied to a step-down substation where it is reduced from 35 or 110 kV to 6-10 kV (Fig. 12-1). Then, power is distributed among the substations in the factory's departments and shops. These substations are equipped with 6-10 kV switchgear, to which are connected cable lines energizing high-voltage motors and transformers which step it down to a voltage of 0.4 to 0.23 kV. From the transformers the power is distributed over internal networks of the respective department or shop, operating at a rated voltage of 380 V or 220 V (these may be electric motors, lighting, and the like).

When a factory is supplied with a voltage of 6 to 10 kV, power is fed to the factory's *main distribution centre*, *MDC* (Fig. 12-2), over one or two cable or (although seldom) overhead lines; the number of lines at the same voltage, taking power from the main distribution centre to the transformer substations in the individual departments and shops is usually greater.

* These figures apply to Soviet practice.— *Translator's note.*

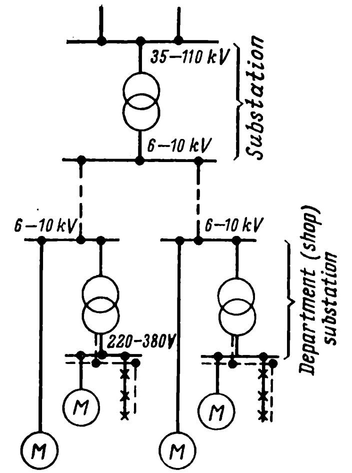


Fig. 12-1. Power distribution at a large factory

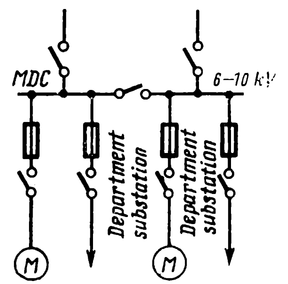


Fig. 12-2. Power distribution through a factory's distribution centre

Medium factories which have no H.V. loads take their power from power-system step-down substations at 6 to 10 kV directly to their own transformer substations from which the voltage stepped down to 0.4-0.23 kV is fed to the distribution centres, *DC*, in the departments and shops.

At small factories with a low installed capacity, power from a power system is taken at 0.4-0.23 kV to a low-voltage distribution centre, whence it is distributed to the various departments and shops*.

At factories, power is distributed from the main distribution centre or a transformer substation to the distribution centres in the departments and shops either over a *radial network* (Fig. 12-3) or a *loop network* (Fig. 12-4).

Among the disadvantages of a radial network distribution system are high cost and reduced security of service, because a fault in the distribution system disrupts all service. On the merit side, it should be noted that a radial

* The rated terminal voltage of step-down transformers must be 0.4-0.23 kV, so that allowing for the voltage loss in the distribution network, the load terminal voltage is 0.38-0.22 kV.

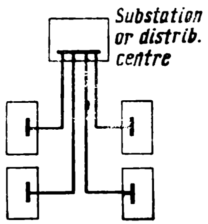


Fig. 12-3. Radial network distribution system

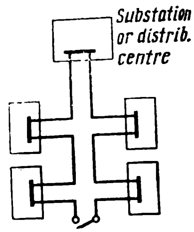


Fig. 12-4. Loop network distribution system

network is simple to operate, protect and control automatically.

Among the merits of a loop network distribution system are low cost and improved security of service, so long as the loop is held completed. Once the loop is opened, the security of service is low.

At factories with critical loads, the security of service is usually enhanced through the use of a loop network energized at two ends from different transformers (or substations).

To minimize voltage variations in luminaires that might be caused by heavy starting currents in induction motors, the lines and networks are sometimes separated into power and lighting.

At a rated supply-line voltage of 380/220 V, motors are connected to the line wires (to obtain 380 V), and lighting fixtures are connected between the neutral and line wires (to obtain 220 V).

Shops with a small number of high-power motors use a radial network distribution system (see Fig. 12-3).

Departments and shops widely use the radial network distribution system—simple and inexpensive and employing bare steel buses along which loads are connected where necessary.

12-2. Industrial Transformer Substations and Switchgear

Switchgear and transformer substations for 35-110 kV are ordinarily of the outdoor type, with all equipment set up in the open. If, however, the air in the locality is laden

with substances detrimental to electrical equipment, switchgear and substations for the same voltages are of the indoor type, that is, set up inside a building. Outdoor devices are less expensive than indoor types, and take less materials and time for their erection.

Wide use is made of metal-clad switchgear and transformer substations.

Metal-clad switchgear and transformer substations have their component units built into metal cells; this is why they are called metal-clad installations.

Cells for metal-clad switchgear are fabricated and the electrical equipment is built into them at factories. Usually, a range of cell types is manufactured, enabling the user to assemble a switchgear installation or a transformer substation to his specifications. This cuts down both the time and money required for the construction of an electrical installation.

An outdoor industrial transformer substation for, say, 36/6-10 kV service will typically consist of a transformer (see Fig. 9-20) which is connected to a 35-kV overhead power transmission line via fuses and disconnecting (isolating) switches. The transformer secondary is connected to the buses of an outdoor metal-clad switchgear installation.

The circuit and a sectional view of a metal-clad switchgear cell are shown in Fig. 12-5. From buses, *1*, current flows through a bus isolator, *2*, an oil circuit-breaker, *3*, a current transformer, *4*, a line disconnecting switch, *5*, to the bushing of a cable or overhead line which delivers it to the buses of a department's or shop's substation.

A factory's main distribution centre which takes power from a power transmission system at 6 to 10 kV is usually set up in a single-storey building and consists either of a metal-clad switchgear cell (see Fig. 12-5) or a cubicle assembled from prefabricated metalwork.

As is with the main distribution centre, the substation in a department or a shop uses cubicles assembled from prefabricated metalwork or factory-assembled and wired metal-clad transformer cells.

The arrangement of equipment at a department (shop) substation is shown in Fig. 12-6. A transformer for a rated voltage of 6/0.4-0.23 kV and a distribution power board

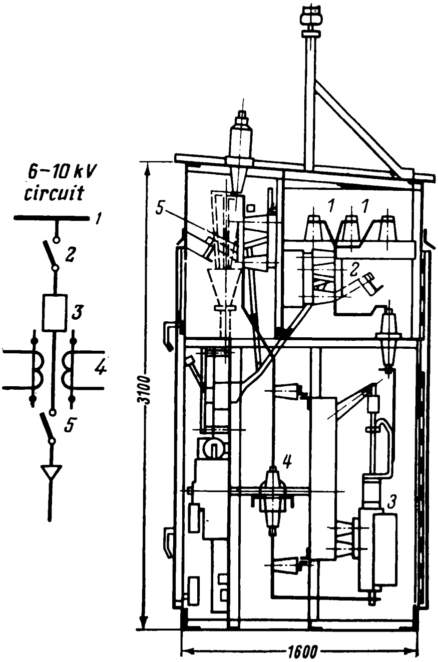


Fig. 12-5. Outdoor metal-clad switchgear cell enclosing an oil circuit breaker and a cable or overhead-line bushing

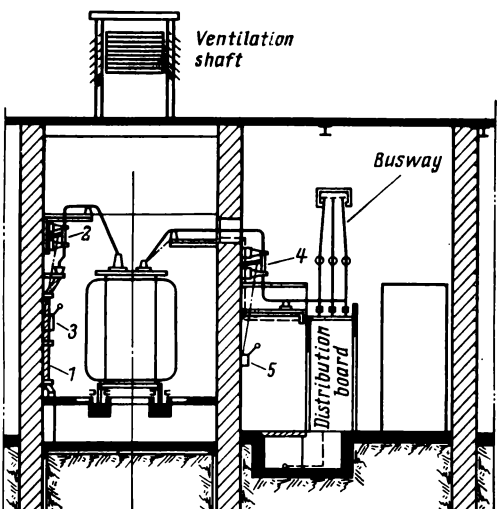
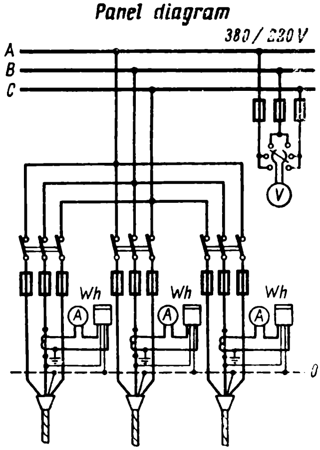
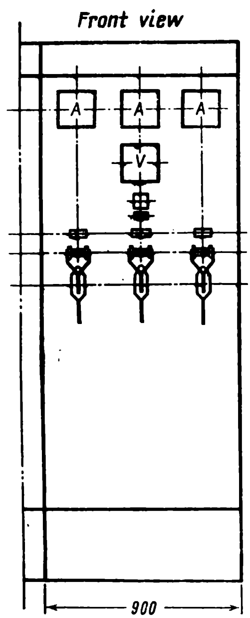
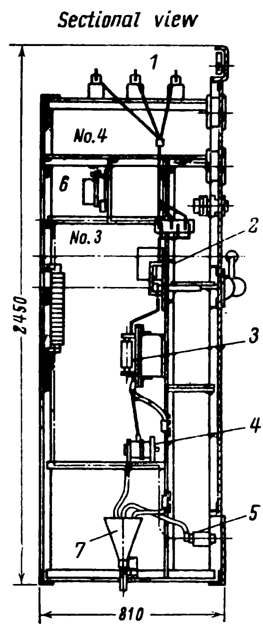


Fig. 12-6. Department (shop) transformer substation



(a)





(b)

Fig. 12-7. Panel of a 380/220 V switchboard and its circuit

(a) diagram; (b) front and sectional views;
1—busbars; 2—knife-blade switch; 3—fuse; 4—current transformer; 5—neutral bus; 6—watthour-meter; 7—cable pot

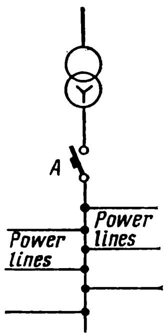


Fig. 12-8. Transformer-loop package

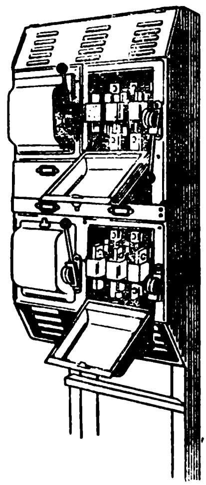


Fig. 12-9. Packaged four-unit distribution board

are installed in adjacent rooms. Power is conveyed to the transformer by cable, *1*, via a lever-operated disconnecting switch, *2*. The secondary leads of the transformer are brought out via a disconnecting switch, *4*, with a lever operating mechanism, *5*, to the buses of the distribution board. Also connected to the same buses are cable lines to energize the lighting circuits and a bus to feed electric motors. Each line has a fuse and a knife-blade switch, and the motor bus circuit is protected by an automatic circuit-breaker. The motor bus way is brought out through a hole in the wall to the adjacent production area.

The low-voltage distribution boards for department substations may be either front-serviced or front-and-back-serviced and of either frame or frameless construction. Frame-type boards use panels from steel sheets or sheets of some insulating material, for example, asbestos-cement. Frameless boards use only sheet-steel panels.

A frame-type front-and-back-serviced board and the circuit of a panel are shown in Fig. 12-7.

To cut down the cost of equipment, use is made of a simplified type of department substation which has no distri-

bution board (Fig. 12-8). Instead, current from the transformer flows via an automatic circuit-breaker directly to a bus whence it is distributed among the individual loads (motors).

Departments with small groups of medium-size loads use distribution centres (Fig. 12-9) assembled from individual cells or boxes enclosing knife-blade or other types of switches and fuses.

12-3. Industrial Power Networks

'a) Overhead and Cable Power Networks

Electric power is transmitted by a system of wires which is called a *power line* if it has no parallel branches, and a *power network*, when it does have parallel branches.

According to the Soviet code for power installations, power lines and networks are classed into those for voltages up to 1000 V and into those for over 1000 V.

According to the function(s) they perform, power lines and networks are further classed into *transmission* ones which convey power from a station or substation to distribution centres, and into *distribution* ones which convey power from distribution centres to loads.

A further classification of networks is into *overhead*, *cable* (or *underground*), and *internal*.

Overhead lines are less expensive to build than underground lines, and they are simpler and more convenient to operate and service because any fault will readily be detected during an inspection. Unfortunately, they are more dangerous and less reliable than underground lines. Overhead lines are predominantly erected on open land and in sparcely populated localities.

An overhead power line consists of *line conductors*, *insulators* and *supports*. The conductors are fastened on the insulators, and these are attached to the supports.

Overhead lines for a voltage of 35 kV and higher use bare aluminium or steel-reinforced aluminium conductors, strings of suspension insulators (Fig. 12-10), and metal, reinforced-concrete or wooden supports (Fig. 12-11). The spacing between the line conductors must be sufficiently

large, so that the air gap between them will not break down even when the conductors are swinging in the wind.

Lines for a voltage of 35 kV and lower often use pin insulators (Fig. 12-12).

Figure 12-13 shows a service entrance for an overhead line, using a porcelain bushing.

Overhead power lines for voltages not over 1000 V mainly use aluminium or steel-reinforced aluminium conductors (for example, a steel-reinforced aluminium conductor may consist of six aluminium wires each 1.8 mm in diameter and one steel wire of the same diameter).

The conductors are installed on low-voltage porcelain insulators (Fig. 12-14) screwed onto pins or the vertical part of hooks with which the insulators are mounted on supports. To meet the requirements for mechanical strength, aluminium conductors must have a cross-sectional area of at least 16 mm^2. The conductors are attached to the neck or (although seldom) head of insulators by soft galvanized steel wire about 1 mm in diameter.

Lines for voltages not over 1000 V use reinforced-concrete or wooden supports about 9 m tall. A wooden support (or pole, as it is usually called) may be in one or two pieces. In the latter case, the lower end of a support is made fast to a reinforced-concrete or wooden stub pole (Fig. 12-15) by a binding of six to eight turns of steel wire about 4 mm in diameter. Supports are erected at intervals of 30 to 80 m, and the conductors are strung so that they are at least 5 m above the ground, at least 20 cm from one another in a horizontal plane, and at least 40 cm in a vertical plane.

Underground lines are used in urban areas and within the premises of industrial enterprises. They use cables. A cable consists of conductors, an insulation, a hermetic sheath or jacket, and an outer covering. The number of conductors varies from one cable to another, and the conductors can be fabricated from copper or aluminium, round or sectoral in cross-section. Each conductor may be from 1 to 240 mm^2 in cross-section, the smaller values applying to cables for voltages not over 1000 V and the larger values, to voltages from 1000 V to 35 kV.

The conductors of a cable are insulated with cable paper impregnated by an oil-resin compound, rubber or poly-

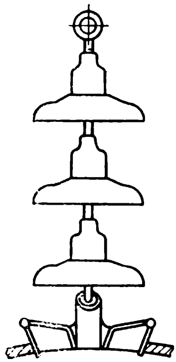


Fig. 12-10. String of suspension insulators

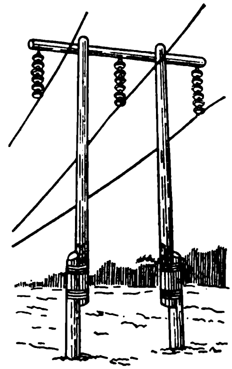


Fig. 12-11. Wooden pole for a three-phase H. V. power transmission line

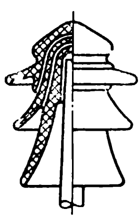


Fig. 12-12. Pin-type insulator for 35 kV

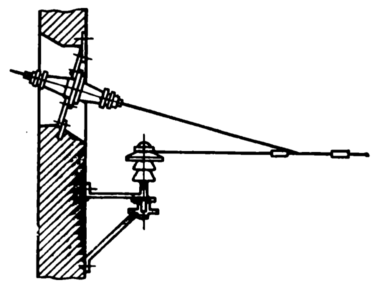


Fig. 12-13. Service entrance for a line

(a) (b)

Fig. 12-14. Attachment of a conductor to a low-voltage porcelain insulator (*a*) and attachment of the insulator on a hook (*b*)

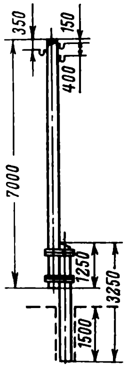


Fig. 12-15. Wooden pole with a stub for a line limited to 1000 V

ethylene. The hermetic sheath or jacket fabricated from lead, aluminium, plastic or rubber is intended to protect the cable against ingress of moisture. Protection against mechanical damage is provided by an armour of two steel tapes or galvanized wire. The armour or sheath is covered by jute fibre impregnated by asphalt compound as a protection against attack by chemicals.

It will be useful for the reader to know how cables are marked in the Soviet Union. Each type designation includes reference to the material of conductors, sheath and covering. For example, the letter A at the beginning of a type designation stands for aluminium conductors. No letter symbol is used in the case of copper conductors. The second letter refers to the material of the protective sheath. For example, the Russian letter "С" stands for "lead-covered" and the second letter "А" for "aluminium-covered". The letter "Б" in the third position may designate an armour or asphalt covering. For example, the designation "ААБ" identifies an aluminium-conductor, aluminium-covered, asphalt covered cable. The designation "АСБ" identifies an aluminium-conductor, lead-covered, steel-armoured cable, jute-and-asphalt covered. An АСБГ cable differs from the last one in that it has no jute-and-asphalt insulation over its steel armour (this fact is indicated by the Russian letter "Г" for "bare").

Figure 12-16 shows the construction of a three-conductor cable for installation in cable tunnels and ducts inside a building. Under the ground, cables are laid in trenches 25

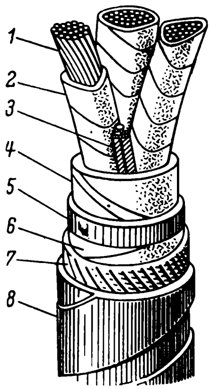


Fig. 12-16. Type СБГ three-conductor cable

1—cable conductors; 2—paper insulation over conductors; 3—fillers; 4—paper belt insulation; 5—lead sheath; 6—paper-asphalt covering; 7—cable yarn; 8—armour of two steel tapes

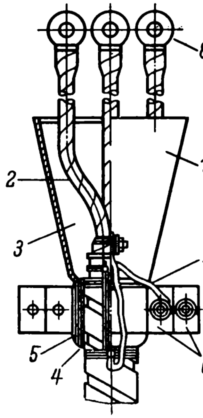


Fig. 12-17. Termination of a three-conductor cable in a steel cone

1—steel cone; 2—insulated conductor; 3—asphalt compound; 4—impregnated-tape covering; 5—impregnated-tape serving; 6—half-clip for attachment; 7—ground wire; 8—cable lug

to 50 cm wide and 70 to 80 cm deep from ground level, or in underground headers.

At the ends, cables are joined by means of lead or cast-iron couplers filled with a compound. The ends of cables connected to machines, transformers or other apparatus are terminated in cable cones (Fig. 12-17) filled with a compound, or they may be terminated "dry", without couplers.

(b) Indoor Wiring

Indoor wiring uses conductors, cords and buses.

A *conductor* is a wire or combination of wires not insulated from one another, suitable for carrying a single electric current. It can be insulated by rubber, PVC or impregnated cotton braiding.

A *cord* is a combination of two or more very flexible conductors insulated from one another.

A *busbar* is a rolled conductor of copper, aluminium or steel, flat or (although seldom) round in section.

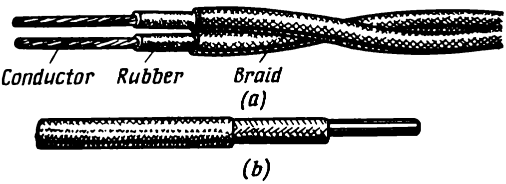


Fig. 12-18. (*a*) ПРД cord; (*b*) ПР or АПР conductor

Conductors, cords and cables may be single- or multi-stranded, and strands may be single- or multi-wire types.

In the Soviet Union, the standard sizes for conductor and cable strands are 0.5, 0.75, 1.0, 1.5, 2.5, 4.0, 6.0, 10, 16, 25, 35, 50, 70, 95, 120, 150, 185, 240, 400, 500, 625 and 800 mm².

In Soviet practice, the following types of conductors, cords and cables are used for lighting and power circuits:

(1) ПРД: flexible, two-wire, cotton-braided, rubber-insulated copper conductor (Fig. 12-18*a*). Comes in sizes from 0.75 to 6 mm² in cross-sectional area, limited to 380 V.

(2) ПР and АПР: a single-wire, impregnated cotton-yarn-braided, rubber-insulated conductor of copper (in the former case) or aluminium (in the latter case) (Fig. 12-18*b*). Come in cross-sectional sizes from 0.75 (ПР) or 2.5 (АПР) to 400 mm², limited to 660 V.

(3) ПРГ: the same as type ПР, but a more flexible stranded-wire conductor.

(4) ПВ and АПВ: differ from types ПР and АПР solely in that they have PVC jackets. Come in cross-sectional sizes from 0.75 to 95 mm²; limited to 660 V.

(5) ПГВ: a flexible, PVC-jacketed conductor. Comes in the same cross-sectional sizes as type ПВ, limited to 660 V.

(6) ШР: is a two-conductor, rubber-insulated cord for the same applications as type ПРД, but has a greater flexibility (Fig. 12-18*a*). Comes in sizes from 0.5 to 1.5 mm²; limited to 220 V.

(7) ППВ and АППВ: flat, two- or three-strand copper or aluminium conductors in PVC jackets. Come in sizes from 0.75 to 2.5 mm² (ППВ) and from 2.5 to 4 mm² (АППВ); limited to 660 V (Fig. 12-19).

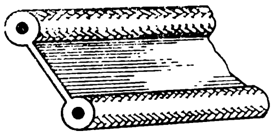


Fig. 12-19. Flat two-conductor wire, ППВ or АППВ

(8) ВРГ and ВРБ: a rubber-insulated, PVC-jacketed cable, with steel-tape armour in the latter case. May have two or four conductors with a cross-sectional area of 1 to 185 mm^2. Limited to 500 V.

(9) НРГ and АНРГ: a plastic-jacketed, rubber-insulated cable with copper (НРГ) or aluminium (АНРГ) conductors. Made with one, two, three or four conductors each with a cross-sectional area of 4 to 185 mm^2; limited to 500 V.

The applications and installation methods for basic types of conductors and cables are listed in Table 12-1.

Copper and aluminium conductors (separate or in cables) can be terminated and joined by compression, welding or soldering (brazing).

When a conductor is terminated in a lug or is joined to the end of another conductor by a compression joint, its end is inserted into the tubular part of a lug in the former case, or the ends of the two conductors are enclosed in a sleeve (connector) of the same material in the latter, and the joint is completed by applying pressure to the sleeve or lug with a hand compression tool or a hydraulic press.

Electric welding is mainly used to joint or terminate aluminium conductors.

Soldering (brazing) is used when compression and welding cannot be applied. Copper conductors are joined by soldering in the flame of a blow-torch, using tin-lead solder and rosin. Aluminium conductors are soldered by zinc-tin solders.

Industrial power circuits are most often wired, using cables, busbars or insulated wires in steel conduits or on insulators.

Industrial lighting circuits have predominantly wires in steel conduits, wires on insulators or knobs (open wiring) or suspended from a steel messenger cable (messenger-cable suspension).

Table 12-1

Applications for Basic Types of Conductors and Cables

Type of wiring	Method of installation	Types of conductors or cables	Location: dry — offices, locker-rooms, etc.	Location: dry — production areas	Damp	Wet	Dust laden	Corrosive air	Outdoor
Open, on insulating supports	On knobs	ПР, ПРД	×	×	—	—	—	—	—
		ПР, АПР, ПВ	×	×	×	×	—	—	×
		АПВ	×	×	×	×	—		×
	On insulators	ПР, АПР, ПВ, АПВ	×	+	+	+	×	+	+
		bare conductors	—	+	×	×	×	×	+
Open, without insulating supports	On surface of walls and ceilings	ТПРФ	×	+	—	—	+	—	—
		ВРГ, СРГ, АСРГ	×	+	+	+	+	×	×
		НРГ, АНРГ	×	+	+	+	+	×	×
		ППВ, АППВ	+	×	×	×	×	×	—
	In metal-covered paper tubing	ПР, АПР, ПВ, АПВ	×	×	—	—	×	—	—
	In steel conduits	ПР, АПР	×	×	×	—	×	—	—
		ПРТО, АПРТО	×	+	×	×	×	×	×
		ПВ, АПВ							

Table 12-1 (cont.)

Type of wiring	Method of installation	Types of conductors or cables	Location						
			dry		Damp	Wet	Dust laden	Corrosive air	Outdoor
			offices, locker-rooms, etc.	production areas					
Concealed	In ducts	ПР, АПР, ПРТО, ПВ, АПВ	—	+	×	—	×	—	—
	In glass tubing	ПР, АПР	+	×	—	—	×	—	—
	In steel conduits	ПР, АПР, ПРТО, АПРТО	× ×	+ ×	+ +	— +	+ +	— ×	— ×
	In semisolid insulating tubing	ПР, АПР	×	×	—	—	×	—	—
	In metal-clad paper tubing	ПР, АПР	×	×	×	—	×	—	—
	In structural elements and under plaster	ППВ, АППВ, ПВ, АПВ, АПН	+	×	×	—	×	—	—

Key: «+» – recommended
«×» – permitted
«–» – forbidden

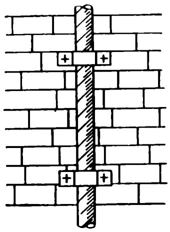


Fig. 12-20. Installation of a cable by means of cleats

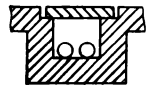


Fig. 12-21. Cables laid in an underground raceway

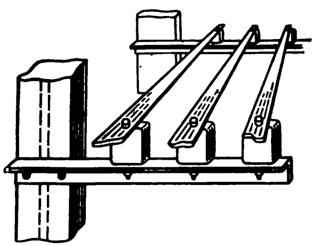


Fig. 12-22. Open busway

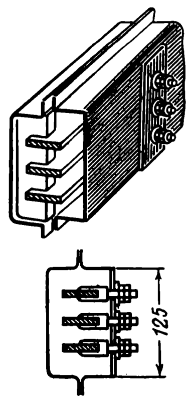


Fig. 12-23. Closed busway

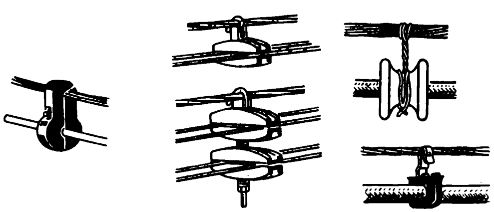


Fig. 12-24. Various modifications of messenger-cable suspension

Cable may be run open or concealed. In the former case, they are installed on walls or ceilings where they are held in place by cleats (Fig. 12-20). In the latter case, they are run in underfloor raceways (Fig. 12-21). Cable termination is discussed in Sec. 12-3*a*.

Busbars can be installed in open and closed bus-ways. *An open bus-way* (Fig. 12-22) is an assemblage of steel structural supports attached to walls, columns, ceilings or joists and fitted with insulators to carry busbars. Branches to loads are made with conductors or cables.

A closed bus-way (Fig. 12-23) is a steel duct enclosing busbars set up on insulating comb-shaped supports. As a rule, they are assembled from standard sections 3 m long. Branches from a closed bus-way are made with suitable tap boxes attached to the main duct. A tap box may enclose either only terminals for branch wires, or cartridge fuses and terminals from which power is taken to loads over insulated wires in steel rigid or flexible conduits. Bus-ways are mounted on steel-tubing supports or brackets attached to supports or columns, or suspended from girders or joists by steel hangers.

In locations with explosion or fire hazard, conductors are run in steel conduits. The conduits must be air-tight, and the connected apparatus and luminaires must be explosion-proof.

In production areas with ramified electrical networks, underfloor conduits are used. Terminations, connections and taps are made with the aid of steel tap boxes.

Installation of insulated wires and cords on knobs and other insulators is now used but seldom.

Industrial lighting circuits are often installed, using messenger-cable suspension. Several modifications of this technique are shown in Fig. 12-24. A messenger-cable suspension span is assembled in advance, complete with all the luminaires and conductors, and then installed in its permanent location by means of support hooks or rod hangers secured at the ends of the span.

The conductors used for messenger-cable suspension are of special types, made integral with an insulated messenger cable. Taps in such cases are made with tap boxes also suspended from the messenger cable.

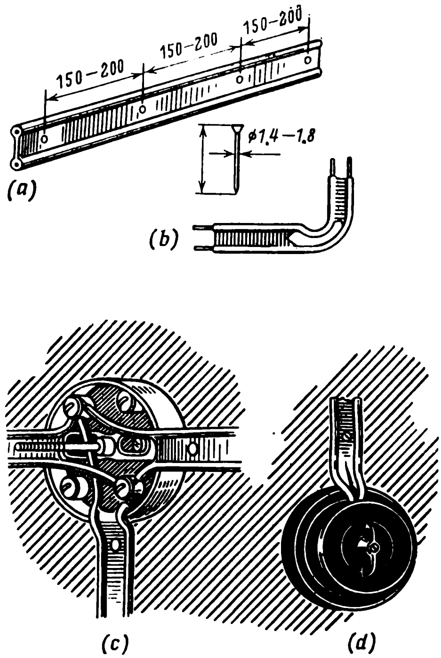


Fig. 12-25. Wiring with ППВ wire
(*a*) straight run; (*b*) curving; (c) tap box; (*d*) connection of a tap to a switch

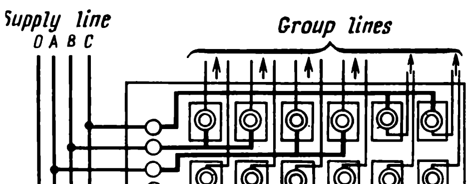


Fig. 12-26. Group lighting board

At present, lighting circuits in residential locations are most often made with flat insulated wires (such as types ППВ and АППВ). They can be run in the open or concealed.

Flat conductors are simple and convenient to install. In the case of concealed wiring, they are laid under plaster, using no additional protective covering. In open wiring, they can be attached to a wall or ceiling by special grades of cement or simply nails. Figure 12-25 shows details of such wiring.

In addition to wires, a lighting circuit includes lamp holders, socket outlets, switches and fuses or automatic circuit breakers to protect the network against short-circuits.

Several fuses mounted on a common baseplate make up a group fuse-box (Fig. 12-26).

(c) Determination of Conductor Cross-Section on the Basis of Maximum Allowable Temperature

In determining the size (cross-sectional area) of a conductor, one needs to know the *power rating* (or rated power) of the associated load P_n, usually stated on its nameplate (hence, another name for it is the nameplate power), the installed load capacity P_i, defined as the sum of the rated powers of all the individual loads, and the design power P_d, that on which all calculations are based. For these powers, there are respective and identically named currents symbolized as I_n, I_i, and I_d.

In practice, it never happens that all loads are turned on at the same time. Also, motors are never carrying a full load all the time. So, in calculations, one proceeds from the design power P_d which is only a fraction of the total installed load capacity.

The ratio of design power to installed load capacity is called the *demand factor*

$$k_d = P_d/P_i \text{ or } k_d = I_d/I_i \qquad (12\text{-}1)$$

The values assigned to the demand factor are unity for external lighting circuits, 0.7 to 0.8 for residential lighting circuits, and 0.7 to 0.9 for industrial networks.

For a lighting load, the design current for a single-phase a.c. circuit or a d.c. circuit is

$$I_d = k_d P_i / V = P_d / V$$

and for a three-phase circuit

$$I_d = k_d P_i / \sqrt{3}\, V = P_d / \sqrt{3}\, V$$

In a cold-metal-working shop the demand factor is unity in the case of one or two motors, 0.8 in the case of four motors, and 0.6 in the case of six (this is power load).

The rated current for d.c. and three-phase motors is respectively given by

$$I_n = P_n / V\eta \quad \text{and} \quad I_n = P_n / \sqrt{3}\, V\eta \cos\varphi \qquad (12\text{-}2)$$

where η is the efficiency of the motor.

The values of η and $\cos\varphi$ for motors are taken from reference sources or handbooks. In approximate calculations for low-power motors (not over 10 to 12 kW), the product of η and $\cos\varphi$ may be taken equal to 0.7-0.8.

Then the design current of motors is

$$I_d = k_d I_n = k_d I_i$$

As a rule (in Soviet practice), the conductor size on the basis of maximum allowable heating is chosen from tabulated data similar to those given in Table 12-2 which specifies the values of I_{max} in continuous duty for the standard sizes of various conductors.

The maximum allowable current for a conductor must be not less than its design current

$$I_{max} \geqslant I_d \qquad (12\text{-}3)$$

Thus, one chooses the wires with a cross-sectional area for which the maximum allowable current is equal to or somewhat greater than its design current.

Example 12-1. Given: A three-phase, loop network operating at 220 V, feeding three motors rated at $P_{n1} = 4.5$ kW, $P_{n2} = 2.8$ kW, and $P_{n3} = 3.5$ kW. To find: The design current; select the size of ПР conductor laid in conduits, on the basis of maximum allowable temperature.

Solution.

Table 12-2

Maximum Allowable Load Currents for Insulated Copper and Aluminium Conductors and Cables

Size, mm²	Maximum allowable load current, A*: open wiring, types ПР, ПРД, ПВ, ППВ, АПР, АПВ	open wiring, copper, types СРБГ, ВРГ, ВРБГ, ТПРФ: 2-conductor	open wiring, copper, types СРБГ, ВРГ, ВРБГ, ТПРФ: 3-conductor	Types ПР, ПРГ, ПВ, ПГВ, АПР and АПВ in single conduit, and ППВ as concealed wiring: 2-conductor	Types ПР, ПРГ, ПВ, ПГВ, АПР and АПВ in single conduit, and ППВ as concealed wiring: 3-conductor	Bare single conductors run outdoors
1	17/—	—	—	16/—	15/—	—
1.5	23/—	19	19	19/—	17/—	—
2.5	30/24	27	25	27/20	25/19	—
4	41/32	38	35	38/28	35/28	50/40
6	50/39	50	42	46/36	42/32	70/55
10	80/55	70	55	70/50	60/47	95/75
16	100/80	90	75	85/60	80/60	130/105
25	140/105	115	95	115/85	100/80	180/135
35	170/130	140	120	135/100	125/95	220/170
50	215/165	175	145	185/140	170/130	270/215
70	270/210	215	180	225/175	210/165	340/265

* The numerator specifies load current for copper and the denominator for aluminium conductors.

The installed capacity is

$$P_i = 4.5 + 2.8 + 3.5 = 10.8 \text{ kW}$$

The design current in the loop network is

$$\begin{aligned} I_d &= k_d P_i \times 1000/1.73 V\eta \cos\varphi \\ &= 0.9 \times 10.8 \times 1000/(1.73 \times 220 \times 0.73) \\ &= 35 \text{ A} \end{aligned}$$

This design current, $I_d = 35$ A, is the same as the maximum allowable load current, $I_{max} = 35$ A, for a ПР wire laid in a conduit (see Table 12-2) with a cross-sectional area (size) of $S = 4$ mm². So we choose this size for the specified service condition.

The conductor size thus chosen must be checked for voltage loss.

(d) Determination of Conductor Size on the Basis of Voltage Loss

As will be recalled (see Sec. 2-10), the voltage loss is defined as the arithmetic difference of the voltages at the start and finish of a line, that is

$$\Delta V = V_1 - V_2$$

Often, the voltage loss is expressed as the percentage of the input voltage and is called the percent voltage loss:

$$\varepsilon = (\Delta V/V) \times 100\% \tag{12-4}$$

The maximum allowable percent voltage loss between a substation and a lighting load is 2 or 3%, and a power load, 4 to 6%.

In Sec. 2-10, we have derived an equation (2-33) which gives the conductor size for a d.c. two-wire line

$$S = 2Il/\gamma\Delta V$$

On replacing ΔV by the percent voltage loss, we get

$$S = 2 \times 100Il/\gamma\varepsilon V$$

or, by multiplying and dividing by V, we may re-write it as

$$S = 2 \times 100Pl/\gamma\varepsilon V^2 \tag{12-5}$$

Hence,

$$\varepsilon = 2 \times 100Pl/\gamma S V^2 \tag{12-6}$$

Equations (12-5) and (12-6) are used to determine the conductor size(s) for a loaded line on the basis of the specified voltage loss or, conversely, the percent voltage loss on the basis of the specified conductor size.

These equations are applicable to d.c., single- and three-phase networks. In the latter case the factor 2 in the numerator must be discarded, V is the line voltage, that is, $V = V_l$, and P is the active power in a three-phase load.

Example 12-2. Given: A three-phase line operating at $V = 220$ V, using type ПР wire 15 m long and 4 mm² in cross-section, feeding three motors rated at $P_1 = 4.5$ kW,

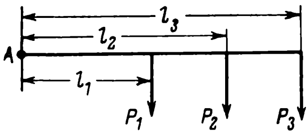


Fig. 12-27. Line feeding three loads

$P_2 = 2.8$ kW and $P_3 = 3.5$ kW (see Example 12-1). To find: The voltage loss.

Solution.

The power in the supply circuits of the three motors operating at rated load is

$$P = P_1/\eta_1 + P_2/\eta_2 + P_3/\eta_3$$
$$= 4.5/0.85 + 2.8/0.85 + 3.5/0.85$$
$$= 5.3 + 3.3 + 4.1 = 12.7 \text{ kW}$$

The design power is

$$P_d = k_d P = 0.9 \times 12.7 = 11.5 \text{ kW}$$

Using Eq. (12-6), we may write

$$\varepsilon = 100 P_d l/\gamma S V^2 = (100 \times 11.5 \times 1000 \times 15)/(57 \times 4 \times \times 220^2) = 2\%$$

Thus, the percent voltage loss does not exceed the allowable limit. So, the conductor size chosen on the basis of maximum allowable heating (maximum allowable load current), $S = 4$ mm^2, is acceptable.

Suppose a line is conveying power from a load centre, A (Fig. 12-27), to several loads at different locations. Then, if all the segments of the line use the same conductor size and material, the conductor size for two-wire d.c. and single-phase a.c. lines may be found by the following equation

$$S = \frac{2 \times 100 (P_1 l_1 + P_2 l_2 + P_3 l_3 + \ldots)}{\gamma \varepsilon V^2}$$
$$= 2 \times 100 \Sigma P l/\gamma \varepsilon V^2 \qquad (12\text{-}7)$$

and the percent voltage loss by the following equation

$$\varepsilon = \frac{2 \times 100 (P_1 l_1 + P_2 l_2 + P_3 l_3 + \ldots)}{\gamma S V^2}$$
$$= 2 \times 100 \Sigma P l/\gamma S V^2 \qquad (12\text{-}8)$$

In the case of three-phase lines, conductor size and percent voltage loss are found by equations which differ from (12-7) and (12-8) only in that there is no factor 2 in their numerators.

The last two equations differ from (12-5) and (12-6) in that the product Pl, called the *load moment*, is replaced by the sum of load moments (see Fig. 12-27).

The conductor size found on the basis of maximum allowable load current and meeting the requirement for allowable voltage loss is then checked for mechanical strength. In our case, this can be done by reference to Table 12-3.

Table 12-3

Minimum Conductor Size Satisfying Mechanical Strength Requirements

Conductors and wiring type	Size, mm^2	
	copper	aluminium
Wires to lighting fixtures	0.5, 1	—
Wires in flexible conduits for mobile loads	1, 2.5	—
Insulated wires indoors, on insulating supports spaced apart by, m:		
up to 1	1.0	2.5
up to 2	1.5	2.5
up to 6	2.5	4
up to 12	4	10
over 12	6	16
Bare conductors indoors, insulated and bare conductors, protected, outdoors:		
on walls	2.5	4
in all other cases	4	10
Insulated conductors in conduits	1	2.5
Overhead lines limited to 1000 V	6	16
Overhead-line drop wires with 25 m to support	4	10

12-4. Safety Grounding

It has been found that, when allowed to pass through a human body, a current of 50 mA is dangerous and a current of 100 mA and higher is lethal to man. It is also known

that the strength of current depends on the voltage across and the resistance of a circuit. So the risk of an electric shock rises with increasing voltage and decreasing resistance of the human body. Damp or wet spots on the skin, contamination of the skin with current-conducting substances and damp or wet footwear increase the risk of an electric shock. Conversely, the resistance of the human body can be enhanced and the risk of an electric shock reduced by wearing rubber gloves and footwear, and using rubber mats.

Lethal outcomes have been recorded at voltages under 60 V. So, the safe voltage for portable lamps and instruments used in a dry room with a wooden floor is assumed to be under 36 V, while for locations with a damp or hot atmosphere, in boilers, tanks, and similar situations the safe voltage is not over 12 V. A very important requirement in all cases is that one of the secondary terminals of the step-down transformer feeding the above loads and its frame must be grounded.

To prevent attending personnel from contact with live parts, use is made of enclosures, fences and interlocks for both the equipment proper and for the locations where the equipment is installed.

To protect attending personnel from an electric shock that might be caused by an inadvertent contact with the metal parts of an electric plant, such as the frame of a motor or of a switchboard, normally supposed to be de-energized, resort is made to *safety grounding*.

Safety grounding is an intentional connection provided between ground and the metal parts of equipment, normally not under voltage. Connection to ground is made by means of *grounding electrodes* which are conductors in direct contact with the soil.

In three-phase networks with the neutral grounded (Fig. 12-28), contact with the ungrounded frame of a motor in which the insulation of a phase is damaged would cause a current of $I_h = V/(r_{ins} + r_h)$ to flow through the human body (the circuit is completed through the insulation resistance of the other phases), which is dangerous, because the insulation resistance may happen to be low.

If, on the other hand, the frame of the motor is grounded (Fig. 12-29), the current that would flow through the dama-

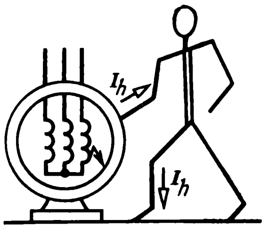


Fig. 12-28. Contact with an ungrounded motor frame

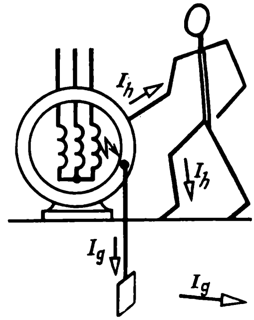


Fig. 12-29. Contact with a grounded motor frame

ged insulation and grounding connector would be

$$I_g = V/(r_{ins} + r_g)$$

and the voltage between the motor frame and ground would be $V_g = I_g r_g$, and it would decrease with decreasing resistance of the grounding. Should a person touch the frame of the motor, he will be connected in parallel with the grounding resistance, so the voltage across him will be low, $V_g \ll V$, and the risk of an electric shock will be avoided.

For example, at $r_g = 4$ ohms and the human body's resistance of 40,000 ohms, the current flowing through the body will be one-ten thousandth of the current flowing through the grounding electrode.

In networks with the neutral wire ungrounded, safety grounding is installed as shown in Fig. 12-30, making sure that the grounding resistance does not exceed 4 ohms; if the supply generator or transformer is rated at over 100 kVA, the grounding resistance must be not over 10 ohms.

Grounding can be obtained by connection to the metalwork of a building, the pieces of equipment in intimate contact with the soil, pipelines (except those carrying gas or imflammable liquids) buried in ground. In the absence of such items, use may be made of steel pipes 2 or 3 m long, 35 to 40 mm in diameter with a wall thickness of at least 3.5 mm, or angle iron with a leg size of at least 4 mm. There should be at least two grounding electrodes driven into

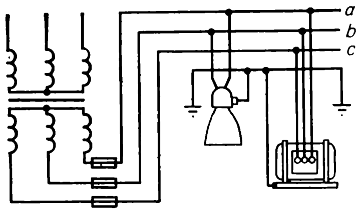


Fig. 12-30. Three-wire system with the neutral wire grounded

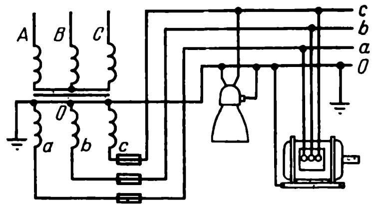


Fig. 12-31. Four-wire system with a solidly grounded neutral wire

the soil so that the upper end of an electrode is 0.4 to 1.5 m below ground level, and interconnected to each other by welded steel strips at least 4 mm thick.

All grounded parts of an electrical installation are connected to grounding electrodes by conductors made of steel strips with a minimum cross-sectional area of 24 mm^2 and a minimum thickness of 3 mm, or steel rounds at least 5 mm in diameter. Grounding conductors are connected to pieces of equipment by welding or bolts, and painted an assigned colour (violet in the Soviet Union).

Electrical installations intended to feed both lighting and power loads ordinarily use four-wire lines operating at 380/220 or 220/127 V. The fourth, or neutral, wire is grounded at both the source and the load (Fig. 12-31) via a very low resistance, so it is, naturally, at ground (or nearly zero) potential. This neutral wire is used for connection of the metal parts of the equipment normally not under voltage. These may be the frames of motors, transformers, and luminaires, the windings of instrument transformers, the metal frames of switchboards, etc.

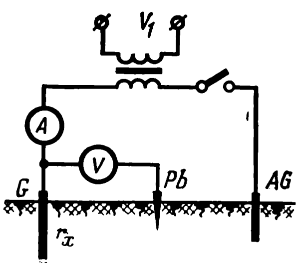


Fig. 12-32. Set-up to measure the grounding resistance by an ammeter and a voltmeter

When one of the phases, for example a is grounded, a fault to ground occurs, and a fuse or an automatic circuit will disconnect the faulty phase from the source. The phase voltages of the phases, V_B and V_C, will remain unaffected.

The state of grounding must be inspected at least once a year and its resistance, $r_g = V/I$, measured using the ammeter-voltmeter method (Fig. 12-32), where G is the grounding electrode under test, AG is an auxiliary ground, and Pb is a probe which is a grounded metal rod for connection of a voltmeter.

Part Two Basic Electronics

Electronics is that field of science and engineering which deals with (1) the electronic and ionic processes taking place in a vacuum, gases, liquids, solids, the plasma, and at their interfaces; (2) the design and properties of electron devices (those in which conduction is by electrons or ions through a vacuum, gas or semiconductor); and (3) the application of such devices, electronic circuits and units in various divisions of science, industry, communication, etc.

To a considerable degree, electronics owes its advances to the progress made by radio engineering. In fact, the two fields have been developing in parallel, because a good many electron devices have served as building blocks for radio equipment and systems. In turn, the problems posed by radio engineering have spurred the search for novel types of electron devices.

In brief, the eventful history of electronics may be summed up as follows. During the period between 1904 and 1913, first the diode, then the triode appeared, and the invention of the triode was followed by that of oscillators. Between 1924 and 1931, the tetrode, pentode, hot- and cold-cathode thyratron made their appearance, and many improvements were made in vacuum tubes and their manufacture. Between 1926 and 1932, first the copper-oxide, then the selenium rectifier was invented.

In the 1930s and later, a good deal of effort was put into semiconductor electronics. In 1948-1951, the point-contact and junction germanium transistors were made, and in 1959-1960, the thyristor was devised.

Naturally, all of these inventions have been accompanied by further advances in their theory and design, including the older vacuum and gas-filled tubes and the newer semiconductor devices.

Until about 1950, electronics had mainly relied on vacuum and gas-filled tubes. Recently, there has been a growing trend towards the use of semiconductor devices as they are smaller in size and weight, draw less power, are more rugged, and more reliable. This does not, however, imply that semiconductor devices will ever oust vacuum and gas-filled tubes completely. There will always be situations where tubes may be preferable much as semiconductor will be in other.

Chapter Thirteen A General Outline of Electronic Processes. Electron Devices

13-1. Classification and Use of Electron Devices

*Vacuum tubes** are defined as ones in which conduction is by electrons moving through a high vacuum, so, they cannot collide with gas molecules. They include, for example, vacuum diodes and triodes, some photocells (or phototubes) and cathode-ray tubes.

Vacuum tubes are used in rectifiers, amplifiers, oscillators, radio receivers, automatic control systems, telemetry and telecontrol, measuring instruments and computers.

Gas-filled tubes are those in which conduction is partly by electrons and partly by ions produced as electrons ionize the filling gas or mercury vapour. The main types of gas-filled tubes are thermionic (or hot-cathode) tubes, thyratrons, and mercury-pool rectifier tubes such as ignitrons.

Because ions are heavy, gas-filled tubes suffer from an appreciable time lag and can only be used in circuits operating at frequencies of several kilohertz, such as medium- and high-power rectifiers, or automatic control circuits.

Solid state electron devices are mainly based on semiconductors in which conduction depends on the movement of electrons (negative charge carriers) and holes (positive charge carriers).

Solid-state electron devices are finding ever increasing use owing to the advantages they offer over vacuum and gas-filled tubes. Their main virtues are as follows: they draw little power, are small in size, light in weight and low in cost, mechanically strong, have a long service life, and are

---

* Usually called "valves" in UK usage.— *Translator's note.*

simple to operate and maintain. In some fields of radio and power engineering, automatic control systems, telemetry and telecontrol, and computers, they vie successfully with vacuum and gas-filled tubes.

13-2. The Motion of Electrons in an Electric Field

Assume that an electron released by the cathode of a tube with initial velocity $v = 0$ is moving in a uniform electric field of strength $\mathscr{E}$ (Fig. 13-1). Then, the force that the field exerts on the electron is

$$F = e\mathscr{E} = eV/d \tag{13-1}$$

The direction of this force is opposite to that of the field because the charge of the electron is negative. The force F imparts the electron an acceleration a which is proportional to F and inversely proportional to electron mass m

$$a = F/m = e\mathscr{E}/m = (e/m)\,(V/d) \tag{13-2}$$

where $e/m = 1.6 \times 10^{-19}$C/$9.1 \times 10^{-31}$ kg = (approx.) 1.76×10^{11} C/kg is the ratio of electron charge to electron mass.

This is an *accelerating field* because the direction of its force F coincides with that of electron velocity v.

Owing to its uniformly accelerated motion, the electron reaches the plate of the tube at a speed v and a kinetic energy

$$W = mv^2/2 \tag{13-3}$$

The electron has gained this energy along the path length d, as the field has done on it the work given by

$$W = Fd = e\mathscr{E}d = eV = mv^2/2 \tag{13-4}$$

Therefore, the energy of the electron is equal to the work done by the field in moving the electron between two electrodes with a potential difference V.

Taking the charge of an electron equal to unity at $V = 1$ V, we shall obtain the unit of electron energy, which is called the electronvolt (1 eV).

Since the charge of an electron is 1.6×10^{-19} C,

$$1\ \text{eV} = 1.6 \times 10^{-19}\ \text{C} \times 1\ \text{V} = 1.6 \times 10^{-19}\ \text{J}$$

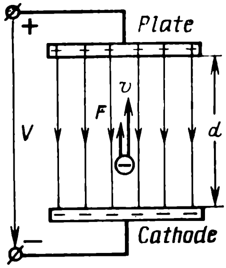


Fig. 13-1. Electron in an accelerating electric field

On the basis of Eq. (13-4), the velocity of an electron in an accelerating field is

$$v = \sqrt{2(e/m)V} = \text{(approx.)}\ 600\sqrt{V}\ \text{km/s} \qquad (13\text{-}5)$$

As is seen, the velocity of an electron depends on the potential difference it falls through. If, for example, an electron leaves the cathode of a tube at v_0 = (approx.) 0, the potential difference being $V = 100$ V, it will reach the plate at $v = 600\sqrt{100}$ = (approx.) 6000 km/s.

The time required for an electron to travel from one electrode to the other can easily be determined provided that we know the distance between the electrodes. For instance, if $d = 2$ cm, then $t = 2d/v_{av}$ = (approx.) $2 \times 2 \times 10^{-5}/6 \times 10^{3}$ = (approx.) 0.7×10^{-8} s.

Referring to Fig. 13-2, let an electron leave the plate with initial velocity $v_0 > 0$ and move to the cathode. The force F that the field exerts on the electron is opposite to the direction of the field and the velocity of the electron, so the field retards the electron, and the electron is under uniform deceleration. In this case, the field is called *retarding*.

During its motion in a retarding field, the kinetic energy $W_0 = mv_0^2/2$ that an electron can have initially decreases because it is expended to overcome the opposition of the field.

If the initial energy of an electron W_0 exceeds that necessary for it to travel between the electrodes, that is, $W_0 > W = eV$, the electron will cover the distance d between the electrodes and reach the cathode. If $W_0 < W = eV$, the electron will have lost all of its energy before striking

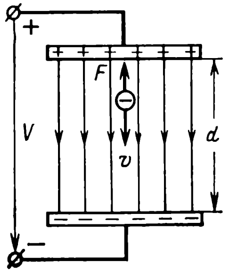


Fig. 13-2. Electron in a retarding electric field

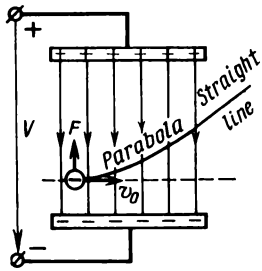


Fig. 13-3. Electron in a transverse electric field

the cathode and it will stop for a very brief moment to be accelerated by the field back to its source electrode. On moving back, the electron will be in a motion with constant acceleration, and the field will give it back the energy the electron lost during its motion to rest.

Now let an electron moving at right angles to an electric field enter it at velocity v_0 (Fig. 13-3). The force F that the field exerts on the electron is, as always, directed opposite to the field. As a result, the electron will simultaneously be in two motions at right angles to each other: the uniform motion due to inertia, perpendicular to the field, and the motion with constant acceleration due to the field but against the field. The resultant path will be a parabola (Fig. 13-3). Should the electron move beyond the field, it will keep moving forward by inertia at constant velocity.

13-3. The Motion of an Electron in a Uniform Magnetic Field

Some electron tubes depend for their operation on the interaction between electrons and a magnetic field.

In Sec. 3-5, we have obtained Equation (3-13) for the force exerted by a uniform magnetic field on an electron moving at right angles to the field. This force is $F = Bev$. We have also found that the direction of this force can be determined by Flemming's left-hand rule.

This expression suggests that when v is zero, F is also zero, that is, the magnetic field does not act on a stationary

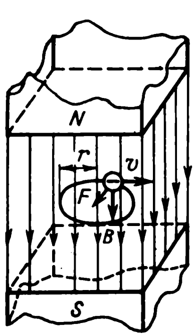


Fig. 13-4. Motion of an electron in a magnetic field, with initial velocity v perpendicular to the magnetic induction vector

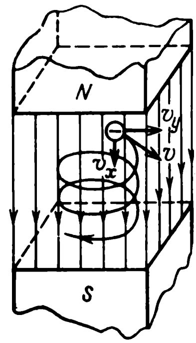


Fig. 13-5. Motion of an electron in a magnetic field, with an initial velocity at an angle to the magnetic induction vector

electron. As the direction of F is at right angles to the electron velocity, it does no work. Thus, the energy and velocity of the electron are not changed in magnitude, but the direction of electron velocity is changed.

If an electron is only affected by a magnetic field, it will describe a circle with a radius r, lying in a plane perpendicular to the direction of the field (Fig. 13-4).

F is a centripetal force given by $F = mv^2/r$.

Therefore, we can write

$$mv^2/r = Bev \tag{13-6}$$

Hence, the radius of the circle

$$r = (m/e)\ (v/B) \tag{13-7}$$

The ratio m/e is constant, so the radius of the circle is proportional to electron velocity and inversely proportional to the magnetic induction.

If an electron enters a field at other than right angles, its initial velocity vector should be resolved into rectangular components. One, v_y, is at right angles to the field, and the other, v_x, is in the direction of the field (Fig. 13-5).

The component v_y results in a circular motion of the electron in a plane at right angles to the field, and the v_x component causes the electron to travel straight on

and at constant velocity in the direction of the field. The path in space which results from such a superposition of velocities is a helix (see Fig. 13-5).

13-4. Electron Emission

As applied to electronics, vacuum is a relative term. If electrons within a device travel practically without colliding with residual gas molecules, we speak of a high, or hard, vacuum. The pressure is about 10^{-5} Pa = = (approx.) 10^{-7} mm Hg.

Vacuum is a nonconducting medium. To maintain an electric current in a vacuum, one needs a source of charged particles (electrons). Usually, such a source is a metal electrode called a *cathode*. Under certain conditions the cathode releases electrons from its surface — a process which is termed *electron emission*.

In the absence of an external electric field, free electrons in metals, known as valence electrons, are free to move at random between the ions of the lattice.

At normal temperature, electrons cannot break loose from the metal, because their kinetic energy is insufficient. Some high-energy electrons, as they move about at random, come up to the surface of the metal, forming an electron layer which, together with the ions of the lattice beneath, forms a dipole (or double-electric) layer (Fig. 13-6). Thus, there is a certain potential difference between the metal and vacuum, which is called a *potential barrier*.

The electric field of this dipole layer counteracts the tendency of electrons to leave the metal, or retards them.

For an electron to escape from the metal, it must acquire an energy equal to the work done in overcoming the potential barrier or the retarding effect of the dipole layer. This energy (work) is called the *work function* of the metal (W_a). The quotient of the work function by the charge of an electron gives the *escape potential* corresponding to the work function

$$\varphi_a = W_a/e$$

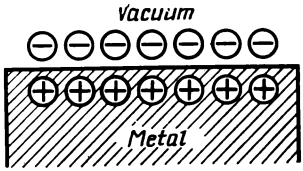


Fig. 13-6. Double-electric layer at the surface of a metal

The work function or escape potential is not the same for different metals. For example, $\varphi_a = 2.1$ V for barium or $\varphi_a = 4.4$ V for tungsten.

Different ways of supplying the escape energy give rise to several forms of electron emission, namely *thermionic emission*, *field emission* (also known as the *autoelectronic effect* or *cold emission*), *photoelectric emission* (or *photoemission*), and *bombardment emission*.

If energy is supplied solely by heating the cathode, the result is known as thermionic emission. Heating raises the velocity and kinetic energy of electrons, so that the number of electrons leaving the metal increases. All electrons escaping from the cathode per unit time, if they are swept off by an external field, form the *emission current* I_e. The emission current rises with increasing temperature.

If there is no accelerating field to sweep them off, the escaping (or emitted) electrons gather in front of the cathode to form a *space charge*. The space charge establishes a retarding electric field near the cathode, which impedes further escape of electrons from the cathode.

If the escape energy is supplied by a very strong electric field, the autoelectronic effect (or field emission) results.

The force acting on an electron in an electric field is proportional to the charge of the electron and field strength, $F = e\mathscr{E}$. When a sufficiently strong accelerating field is applied, the forces exerted on the electrons near the cathode surface are great enough to overcome the potential barrier and pull the electrons from the cold cathode.

It is also possible to supply the escape energy by illuminating the cathode surface with no heating of the cathode. The electrons acquire the energy necessary for escape as the cathode absorbs the incident particles of light, photons.

Radiant energy is emitted and absorbed in portions known as *quanta*. If the energy of a quantum, which is defined as $W_q = h\nu$, where h is Planck's constant and ν is the frequency of the radiation, exceeds the work function W_a for the material of the cathode, the electron may leave the cathode, in which case, photoemission takes place.

In *secondary emission*, electrons, called primary, strike the surface of a conductor or semiconductor; the collision results in releasing more electrons which are called secondary. On striking the conductor, electrons penetrate its surface layer and give up some of their energy to the electrons of the conductor. If, as a result, the energy of the secondary electrons exceeds the work function, they will leave the conductor.

A primary electron having a considerable energy may give it up to either one or more electrons, so the number of secondary electrons may be greater than that of primary ones.

In bombardment emission, the cathode is struck by heavy ions or excited atoms (molecules), with the result that electrons are knocked out of the material. This type of emission resembles closely secondary emission.

13-5. Vacuum-Tube Cathodes

A vacuum-tube *cathode* is that electrode which emits electrons.

Most commonly, use is made of the *thermionic* type, that is, cathodes which function primarily by thermionic emission. They are divided into *directly heated cathodes*, usually in the form of a wire made of a refractory metal heated by filament current (Fig. 13-7*a*), and *indirectly heated cathodes* (also called heater cathodes) made in the form of a nickel cylinder coated with a layer of emitting material, which is slipped over an insulated heater carrying filament current (Fig. 13-7*b*).

The main characteristics and ratings of cathodes are given below.

1. *The emissive capacity of a cathode* is defined as the unit thermionic emission or electron emission current

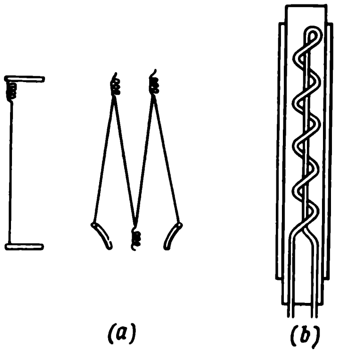


Fig. 13-7. Types of cathode construction

(*a*) directly heated; (*b*) indirectly heated

density $j_e = I_e/S$ at the nominal temperature of the cathode, where S is the surface area of the cathode. Its value is about hundreds of milliamperes per square centimetre.

2. *The unit filament dissipation* is the filament dissipation of a cathode per square centimetre of its surface area, that is

$$P'_f = P_f/S = V_f I_f/S \tag{13-8}$$

As little as 2 or 3 per cent of the power expended to heat the cathode is turned into the kinetic energy of electrons leaving the cathode, the remainder being lost to the surroundings by radiation and thermal conduction.

3. *The emission efficiency* is the ratio of the emission current to the filament dissipation

$$H = I_e/P_f = I_e/V_f I_f \tag{13-9}$$

As is seen, the emission efficiency rises with increasing emission current. For present-day cathodes, it ranges from units to hundreds of milliamperes per watt.

4. *The operating temperature* of cathodes runs from 600°C to 2400°C.

As the operating temperature of a cathode rises, there is an increase in the thermionic emission current density and, as a consequence, in the intensity of emission, but its service life is cut down.

5. The cathode ratings also include the *life of the cathode*, which is determined on the basis of its emission. A cathode is said to "have burned out", if its emission has dropped

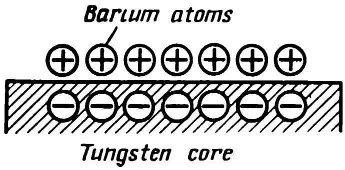


Fig. 13-8. Dielectric layer formed by polarized barium atoms

by twenty per cent. The life of cathodes usually runs into several thousand hours.

The emitters or cathodes as used in present-day thermionic electron tubes may further be divided into three groups, namely *pure metal emitters* (for example, pure tungsten), *atomic-film emitters* (mostly thoriated or bariated tungsten), and *oxide-coated emitters.*

In thoriated (or bariated) tungsten cathodes, the atomic film sets up an accelerating field for the escaping electrons owing to polarization (Fig. 13-8). In oxide-coated cathodes the emitting surface is a mixture of barium, strontium and calcium oxides, with barium atoms dispersed throughout.

Atomic-film and oxide-coated cathodes have reduced work functions, so effective emission can be obtained at a relatively low temperature.

13-6. The Vacuum Diode

(a) Principle of Operation

A vacuum diode is the simplest of all electron tubes. It consists of two electrodes, a plate (or anode) and a cathode, placed in a high vacuum (10^{-5} Pa $\approx 10^{-7}$ mm Hg) metallic, ceramic or glass envelope (Fig. 13-9). The nickel plate P may be in the shape of a cylinder or a box without a top and bottom. The filamentary or indirectly heated cathode K is located inside the plate. The electrodes are often connected to pins press-fitted into a plastic base, or brought out directly through the envelope. The plate is connected to one pin, a filamentary cathode is connected to two pins, and an indirectly heated cathode to three pins. In high-voltage diodes, the plate lead or, rather, cap is usually located at the top of a tube (Fig. 13-9).

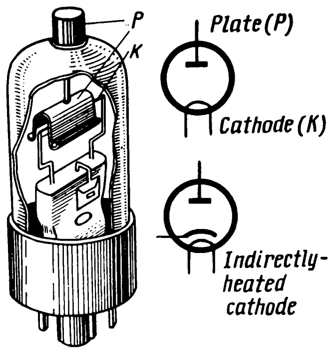


Fig. 13-9. Construction and diagram symbols of a vacuum diode

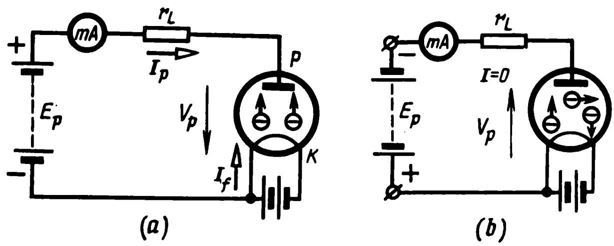


Fig. 13-10. Connection of a vacuum diode in a circuit
(a) with plate voltage applied in the forward direction; (b) with plate voltage applied in the reverse direction

The symbols for diodes with filamentary and indirectly heated cathodes are shown in Fig. 13-9.

The cathode of a tube is fed from a filament-voltage source at about several volts, which may be a battery (usually called the A-battery) or the secondary winding of a transformer (Fig. 13-10). When heated, the cathode emits electrons. To make these electrons reach the plate, there should be an accelerating field between the plate and cathode. For this purpose, the plate is connected to the positive and the cathode to the negative terminal of a plate-voltage source (called the B-battery). The potential difference between the plate and cathode is termed plate voltage V_p. The electrons leaving the cathode and reaching the plate make up plate current I_p. This current is opposite to the direction of the electrons. If we con-

nected the plate to the negative terminal and the cathode to the positive terminal of the B-battery, that is, applied a reverse voltage, the field between the plate and cathode would decelerate the electrons emitted by the cathode, thereby forcing them to return, and no current would be flowing in the plate circuit.

Thus, a diode shows *unidirectional conduction* of current, that is, from the plate to the cathode. This is the main distinction of a diode. Devices conducting current only in one direction are sometimes called *electric valves**.

The electrons escaping from the cathode fill the plate-cathode space. In addition to the accelerating field established by the plate voltage, the electrons closer to the cathode experience the retarding effect produced by the electrons closer to the plate. As a result, the electrons near the cathode form a negative space charge. This space charge opposes the escape of electrons from the cathode and retards their travel towards the plate.

The relation between the plate current and plate voltage at a constant filament voltage, $I_p = f(V_p)$ at $V_f =$ $=$ constant, is called the *plate or volt-ampere characteristic* of the diode. A circuit for measuring this characteristic appears in Fig. 13-11. Figure 13-12 shows two volt-ampere characteristics measured at different filament voltages.

At $V_p = 0$, a very small current, I_{p0}, is flowing through the diode. It is called the *initial diode current*. It is made up of the electrons escaping from the cathode, whose kinetic energy is high enough for them to reach the plate. To reduce this current to zero, it is necessary to establish a retarding field between the electrodes. For this purpose, a reverse plate voltage of about 1 V termed the *cut-off voltage* V_c is applied.

As the positive plate voltage increases, the plate current begins to rise, too, at first slowly, then more rapidly. This is because a rise in the plate voltage reduces the negative space charge which retards the electrons, and boosts the accelerating plate field. This is known as the *space-*

*Most commonly, they are used to rectify alternating current. So, by extension, they are more usually called rectifier tubes.— *Translator's note.*

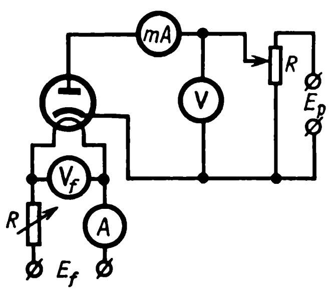


Fig. 13-11. Test set-up for measuring the characteristics of a vacuum diode

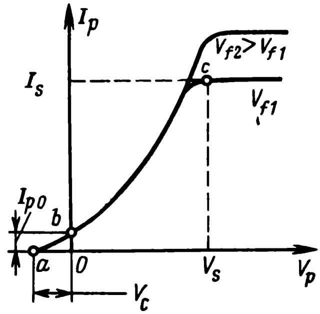


Fig. 13-12. Plate characteristics of a vacuum diode

charge-limited operation of a diode. The condition in which the plate voltage is so high that all electrons escaping from the cathode reach the plate is called *saturation* ($I_p = I_e$) or *temperature-limited operation.* The respective plate current is termed the saturation emission current. A further increase in plate voltage acts on plate current differently. With a tungsten cathode, it remains practically unchanged, with a bariated cathode, it increases slightly, and with an oxide-coated cathode, there is a considerable rise in plate current.

In approximate calculations, the actual characteristic of a diode is often replaced by straight-line segments (Fig. 13-13). This technique is known as *piecewise-linear approximation.*

The analysis of vacuum-diode operation requires knowledge of certain diode factors, generally called *parameters.* Those related to the vacuum diode are plate conductance, dynamic plate resistance, maximum plate dissipation, and peak-inverse voltage.

The *plate conductance* of a vacuum diode, symbolized g_p, is defined as the ratio of an incremental change in plate current ΔI_p to the incremental change in plate voltage ΔV_p that caused it

$$g_p = \Delta I_p / \Delta V_p \tag{13-10}$$

It ranges from 1 to 50 mA/V for different diodes.

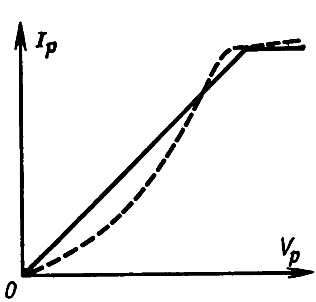


Fig. 13-13. Plate characteristic (the dashed line) of a vacuum diode and its piecewise-linear approximation

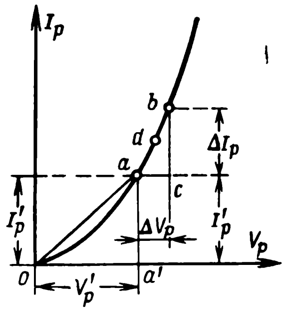


Fig. 13-14. Determining the transconductance and resistances of a vacuum diode

The reciprocal of the plate conductance (or the slope of the plate current/plate voltage curve), $1/g_p$, defined as the rate of change of plate voltage with respect to plate current is termed the *dynamic plate resistance* or the *a. c. resistance* of a vacuum diode:

$$R_p = 1/g_p = \Delta V_p/\Delta I_p \qquad (13\text{-}11)$$

The value of R_p ranges from tens to hundreds of ohms for various types of diodes.

This resistance shows that, when a diode operates on an alternating current, any change in plate voltage causes the respective change in plate current.

In addition to the dynamic plate resistance of a diode, there exists the concept of its *static* or *d.c. resistance* which is defined as the ratio of the d.c. plate voltage V'_p to the respective plate current I'_p (Fig. 13-14):

$$R_0 = V'_p/I'_p \qquad (13\text{-}12)$$

For the most part, the d.c. resistance of a diode R_0 exceeds its a.c. resistance R_p.

As the plate characteristic of a diode is nonlinear, its slope and the a.c. resistance of the diode are not the same at different points on the characteristic.

The plate conductance of a vacuum diode can be determined from its characteristic, as shown in Fig. 13-14, by finding the incremental change in plate current, ΔI_p (por-

tion bc), caused by an incremental change in plate voltage, ΔV_p (portion ac), within the selected portion (ab) of the diode characteristic, and by dividing the former by the latter. Thus, we obtain the plate conductance for portion ab or some point d, midway between points a and b. In some cases, it may be necessary to determine the maximum plate conductance corresponding to the linear part of the diode characteristic.

As they strike the plate, the electrons contributing to plate current give up their kinetic energy, $mv^2/2$, which is converted to heat. Should the power, P_p, delivered to the plate exceed what it gives up (dissipates), the plate temperature would rise. This excess power serves no useful purpose; in fact, it may overheat and damage the plate and destroy the cathode located nearby.

In moving from the cathode to the plate, an electron gains an energy, $mv^2/2 = eV_p$. Assuming that n electrons strike the plate every second, the power delivered to the plate in the same time is

$$P_p = nmv^2/2 = neV_p = I_pV_p \tag{13-13}$$

The limit for the maximum power that can be delivered to the plate by the electrons is set by the maximum allowable plate temperature. This limit is called the *plate dissipation* of a tube.

In operation, plate dissipation P_p must be less than the maximum plate dissipation:

$$P_p = V_pI_p < P_{p,\,\max} = I_{p,\,\max}V_p$$

and the maximum plate current

$$I_{p,\,\max} = P_{p,\,\max}/V_p \tag{13-14}$$

The plates for vacuum tubes are fabricated from nickel, molybdenum, tantalum or graphite.

As a way of raising plate dissipation, measures are taken to cool the plate. For this purpose, the area of the plate is increased by fitting fins or radiators to its surface, or it may be blackened or coated with zirconium to improve radiation and absorb residual gases.

Tube manufacturers specify maximum plate dissipation $P_{p,\,\max}$, at which the plate temperature does not exceed the limit.

The safe value of inverse voltage must always be lower than the peak-inverse voltage. Otherwise, a self-sustaining arc-back may occur, and the insulation may break down. Data given by manufacturers specify the safe value of peak-inverse voltage $V_{p\text{-}i}$. Tube ratings also include nominal (or rated) voltage V_n and nominal (or rated) filament current I_n.

(b) Types of Vacuum Diodes and Their Symbols

According to their purpose, there are two basic groups of vacuum diodes, namely rectifying vacuum diodes and radio-frequency (r.f.) diodes.

As its name implies, a rectifying vacuum diode is intended to rectify power- (or higher) frequency alternating current to direct current.

A radio-frequency (r.f.) vacuum diode is an electron tube designed for signal detection, modulation and frequency conversion at radio frequencies.

Vacuum diodes may have one or two plates. A double diode is a combination of two similar diodes in the same envelope. It can contain one common cathode or two isolated cathodes.

The size of the tube envelope depends on tube power. The point is that bigger tubes dissipate more heat, and it takes a larger surface area to dissipate it.

In addition to conventional base-type glass or metal tubes, there are stiff-lead and soft-lead diodes which have no base (Fig. 13-15). In a stiff-lead tube, the electrode leads also serve as pins. The electrode leads of a soft-lead diode are tinned flexible wires which are directly soldered to appropriate points in a circuit.

For the most part, conventional glass and metal-envelope tubes have an octal (eight-pin) base. The eight pins connected to the electrodes are arranged in a symmetrical pattern round the circumference, with a plastic polarization (also known as alignment or index) pin with a rib in the centre. The alignment pin provides for proper insertion of a tube in its socket (Fig. 13-16*a*). It is customary to number the pins clockwise, beginning from the rib on the alignment pin (Fig. 13-16*c*). Some pins

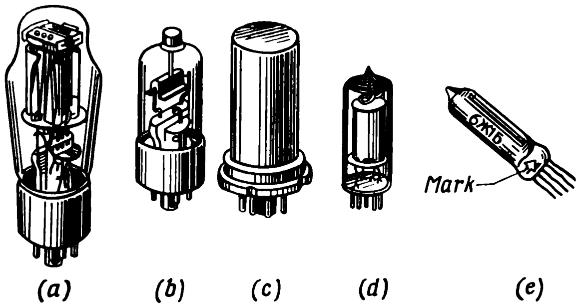


Fig. 13-15. Various types of vacuum tubes
(*a*) a glass-envelope, low-voltage, double-plate rectifying diode; (*b*) a glass-envelope, high-voltage, plate-cap rectifying diode; (*c*) a metal-envelope tube; (*d*) a stiff-lead miniature tube; (*e*) a soft-lead microminiature tube

may be omitted, if no connections need be made to them inside the tube. In stiff- or soft-lead tubes which have no base and with the leads acting as pins, they are numbered, starting from a colour mark on the stem.

The diagram showing the connection of the tube electrodes to the pins is called a base diagram (base diagram can be found in tube manuals).

It will be useful for the reader to know the system adopted in the USSR for marking vacuum tubes. The system established by a USSR State Standard, GOST 13393-67, runs as follows:

The type designation of a tube consists of four elements.

The first element is a number giving a rounded filament voltage (in volts).

The second element is a letter identifying the tube as a diode (Д), double diode (X), or rectifying diode (Ц).

The third element is the ordinal number of a given modification in the type series.

The fourth element is a Russian letter identifying the tube as one in a glass envelope (C) with a diameter of more than 22.5 mm, or in a ceramic envelope (К); or as a miniature glass-envelope tube 19 mm and 22.5 mm in diameter (Г); or as a microminiature tube with a diameter of less than 8 mm (Б) or less than 5 mm (P); or a metal-ceramic tube (H).

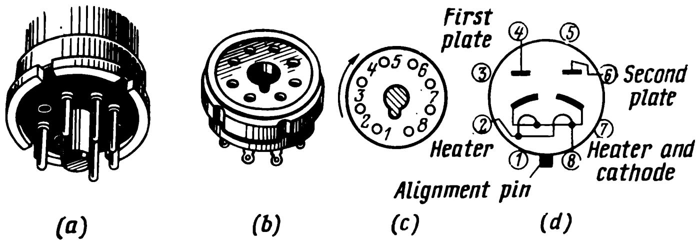


Fig. 13-16. (*a*) An octal base; (*b*) a tube socket; (*c*) pin numbering; (*d*) the basing diagram of a 5Ц4С rectifying diode

The absence of a fourth element is an indication that the tube has a metal envelope.

As an example, here are the type designations of some Soviet-made tubes:

5Ц4С: the letter "Ц" identifies it as rectifying diode, the digit "5" indicates that the filament voltage is about 5 V, the letter "С" refers to a glass envelope, the digit "4" indicates that this is a No. 4 modification tube.

6Д6А: an r.f. diode (Д), with a filament voltage of about 6 V, in a microminiature glass envelope with a diameter less than 8 mm (А), issued as Modification No. 6.

6Х6С: a normal-size, glass-envelope, Modification No. 6, dual diode for a filament voltage of about 6 V (actually 6.3 V).

Chapter Fourteen

Vacuum Triodes and Multi-Electrode Tubes

14-1. The Structure and Principle of Operation of the Vacuum Triode

In contrast to a diode, a vacuum triode has three electrodes, the third one being a *control grid.* Its function is to control the flow of electrons, that is, cathode current of the tube. In most triodes, this electrode is a wire spiral made of tungsten, nickel, molybdenum or their alloys. The grid is located between the plate and cathode, closer to the latter (Fig. 14-1). As in diodes, the cathode of a triode can be either filamentary or indirectly-heated. Triodes using both types of cathodes and their symbols are given in Fig. 14-2.

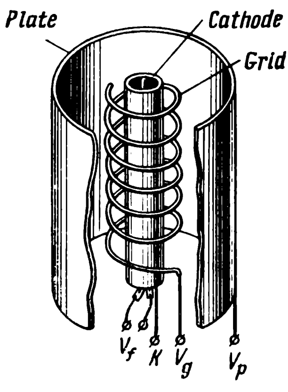


Fig. 14-1. Construction of a vacuum triode

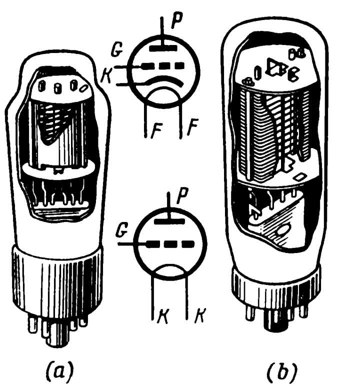


Fig. 14-2. Construction of a triode using an indirectly heated cathode (*a*); a directly heated cathode, and the respective diagram symbols (*b*)

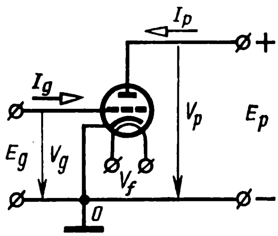


Fig. 14-3. Connection of a vacuum triode in a circuit

In operation, a triode has three circuits (Fig. 14-3), namely, a filament circuit, a plate circuit, and a grid circuit. The last two have a common point O, which is connected to the cathode. This point is usually grounded (connected to the chassis). Its potential is assumed to be zero, so the potentials at any other points in the circuit are stated in relation to this point.

The potential difference between the grid and cathode is known as *grid voltage*.

In a vacuum diode, the electrons escaping from the hot cathode are affected by the space charge and the electric field established by plate voltage V_p. In a triode, the electric field between the plate and cathode is established not only by plate voltage, but also by grid voltage V_g. The effect of grid voltage is greater, because the grid is closer to the cathode. Also, the grid reduces the effect of plate current, for it acts as a screen in the electric field of the plate.

When the grid is made negative, $V_g < 0$, the potentials at points in the field near the grid are reduced. This produces a retarding field in the grid-cathode space, opposing the accelerating field due to the plate. At a constant plate voltage, the resulting field depends on grid voltage. When negative grid voltage is low, the resulting field still maintains its accelerating function, and a plate current I_p will flow between the cathode and plate. When the grid is made sufficiently negative, the resulting field becomes retarding, and no current flows. The point at which this happens is called the *cut-off point*. Accordingly, the tube is said to be cut off, and the voltage responsible for the effect is termed the *cut-off voltage* V_{gc}.

When the grid is made positive with respect to the cathode, $V_g > 0$, it sets up an accelerating field which combines with that due to the plate. As a result, the field strength in the grid-cathode space rises. Plate current will naturally rise, too. At $V_g > 0$, some of the emitted electrons strike the grid, and the tube is then said to draw a *grid current* I_g.

It follows from the foregoing that a change in grid voltage has a greater effect on the field in the grid-cathode space, the space charge and plate current. Thus, the *grid is much more effective in control of plate current than the plate.*

If weak electric signals or oscillations of any waveform or frequency are applied to the grid circuit, they will also appear in the plate circuit or in the load connected in this circuit but with increased power. This is signal amplification which is the principal function of a vacuum triode.

14-2. Static Characteristic Curves of the Vacuum Triode

The performance of a vacuum triode can be presented graphically by static characteristic curves (or simply static characteristics), in which the plate current is related to the plate and grid voltage, with the filament voltage held constant. The most important of them are:

(1) the *plate characteristic* showing the relationship between the plate current and plate voltage, with the grid voltage held constant, $I_p = f(V_p)$ at $V_g = \text{const}$;

(2) the *grid-plate characteristic* representing the relationship between the plate current and grid voltage, with the plate voltage held constant, $I_p = f(V_g)$ at $V_p = \text{const}$.

These characteristics are termed *static* because they are measured when the voltage at one of the electrodes is held constant.

Figure 14-4 shows a circuit arrangement which can be used to measure the static characteristics of a triode experimentally.

The plate characteristic curve of a vacuum triode is plotted as follows. To begin with, the nominal filament voltage and a certain grid voltage are set and maintained

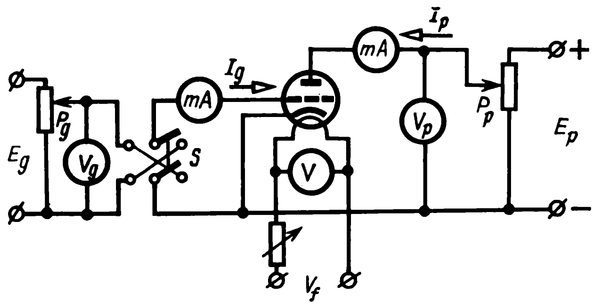


Fig. 14-4. Test set-up to measure the static characteristics of a triode

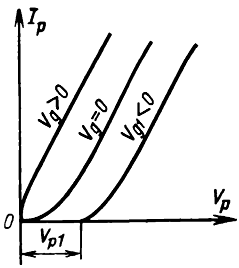


Fig. 14-5. Plate characteristics of a vacuum triode for various values of grid voltage

constant. Then, the plate voltage is raised in equal increments starting from zero, while noting the plate current and the respective plate voltage. Now, the plate characteristic curve can be plotted, for which purpose the plate current values are laid off as ordinates and the values of plate voltage as abscissas. At zero grid voltage, the plate characteristic of a triode does not practically differ from that of a vacuum diode (Fig. 14-5).

The plate characteristic measured at a constant grid voltage V_{g1} will retain its shape and slope, but it will be shifted to the right of the characteristic obtained at $V_g = 0$, that is it will extend into a region of higher plate voltage values. This is because at low plate voltage, the tube is cut off by the negative grid voltage and can be driven conducting only when plate voltage reaches a positive value V_{p1}, such that the retarding field of the grid is fully balanced by the accelerating field of the plate. At a still higher negative grid voltage, the tube will be rendered

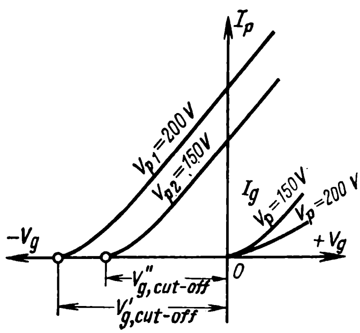


Fig. 14-6. Plate-grid and grid current characteristics of a vacuum triode

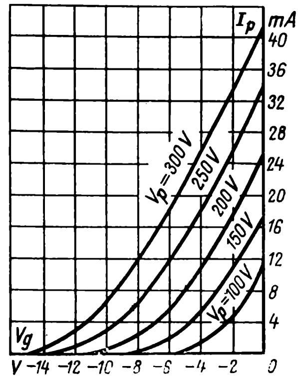


Fig. 14-7. Family of plate-grid characteristics of a vacuum triode

conducting at a higher plate voltage. If the grid is made positive, the curve will shift to the left of the first one, and it will run higher in proportion to the increase in grid voltage. Together, these plate characteristics make a family of plate characteristics.

To measure the grid-plate characteristic of a triode (Fig. 14-6), it is necessary to set the nominal filament voltage, the required plate voltage, and the maximum negative grid voltage. After that, the grid voltage is slowly reduced to zero and the respective values of plate current are noted. Then, the polarity of grid voltage is reversed (with the switch *S* shown in Fig. 14-4) and the part of the grid-plate characteristic corresponding to the positive values of grid voltage is plotted. The characteristic curves for other values of plate voltage can be plotted in the same way. The origin of the grid-plate characteristic corresponds to that value of grid voltage at which the accelerating field in the cathode-grid space due to the plate voltage fully compensates the retarding field due to the negative grid voltage. It is natural that a higher plate voltage results in a grid-plate characteristic having a higher initial value of negative grid voltage. As the grid voltage is decreased, the accelerating field in the grid-

cathode space and, as a consequence, the plate current are increased.

When the grid is made positive, it starts drawing current. In most cases, the grid current is undesirable, because it not only serves no useful purpose, but also has an adverse effect on triode operation. Figure 14-6 shows two grid characteristics, $I_g = f(V_g)$. As the plate voltage rises, the grid current decreases, because in this case the accelerating field near the grid intercepts the electrons at the grid surface, so the likely grid current goes down.

Figure 14-7 shows a family of grid-plate characteristics for different plate voltages.

A tube with sparsely spaced grid wires has a high cut-off voltage, and the characteristic shows a long tail. This is what is called a remote cut-off characteristic (Fig. 14-7), and such a tube is classed as a remote cut-off tube. On the other hand, a tube with closely spaced grid wires has a low cut-off voltage, and the tail of its characteristic is very short. This is called a sharp cut-off characteristic, and such a tube is a sharp cut-off tube.

14-3. Parameters and Ratings of the Vacuum Triode

The parameters chosen for a vacuum triode establish the relationship between plate current, plate voltage and grid voltage.

The main parameters of a triode are its *transconductance* (also known as mutual conductance) g_m, *dynamic plate (a.c.) resistance* R_p, and *amplification factor* μ.

The transconductance of a vacuum triode is the ratio of an incremental change in plate current ΔI_p to the incremental change in grid voltage ΔV_g that caused it, the plate voltage being held constant while the changes take place (Fig. 14-8):

$$g_m = \Delta I_p / \Delta V_g \text{ at } V_p \text{ held constant} \qquad (14\text{-}1)$$

The grid-plate characteristic given in Fig. 14-8 shows that a change of $\Delta V_g = V_g'' - V_g'$ in grid voltage causes a change of $\Delta I_p = I_p'' - I_p'$ in plate current.

The transconductance tells how many amperes should be added to or subtracted from the plate current as the

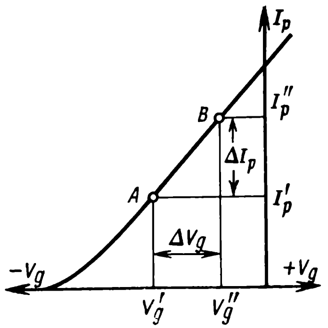


Fig. 14-8. Determining the transconductance of a vacuum triode (the slope of its plate-grid characteristic)

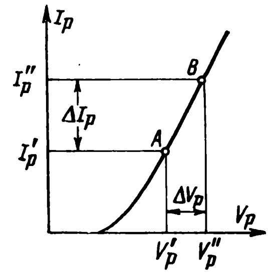


Fig. 14-9. Determining the dynamic plate resistance of a vacuum triode

grid voltage (or bias) is changed by one volt so as to keep the plate voltage constant.

The transconductance is not the same within different portions of the characteristic. In the rising portion, it is higher than in the initial portion. The transconductance of various types of triodes ranges from 1 to 40 mA/V.

Thus, the transconductance shows numerically how effective the grid of a vacuum triode is in controlling the plate current.

The dynamic plate resistance of a vacuum triode is the ratio of an incremental change in plate voltage ΔV_p to the incremental change in plate current ΔI_p which the change of voltage produces, with grid voltage held constant (Fig. 14-9):

$$R_p = \Delta V_p/\Delta I_p, \quad V_g \text{ remaining constant} \qquad (14\text{-}2)$$

The plate characteristic shown in Fig. 14-9 shows that a change of $\Delta V_p = V_p'' - V_p'$ in plate voltage causes a change of $\Delta I_p = I_p'' - I_p'$ in plate current.

The dynamic plate resistance is the internal opposition offered between the cathode and plate of a triode to changes in plate current (as the resistance rises, the plate has a progressively lesser effect on the plate current).

The dynamic plate resistance varies within different portions of the characteristics. In the rising portion, it is less than in the initial portion, and is almost constant. The dynamic plate resistance ranges from 1 to 100 kilohms.

The amplification factor is the ratio of a small change in plate voltage ΔV_p to the small change in grid voltage ΔV_g required to produce the same change in plate current:

$$\mu = -(\Delta V_p/\Delta V_g) \text{ at } I_p \text{ held constant} \tag{14-3}$$

The "minus" sign shows that in order to hold the plate current at a constant value, a change in the plate voltage should be compensated for by an opposite change in the grid voltage or vice versa.

Sometimes, the amplification factor may be defined as the absolute value of the ratio given in Eq. (14-3):

$$\mu = |\Delta V_p|/|\Delta V_g|$$

The amplification factor tells how many times more effective the grid voltage is than the plate voltage in controlling the plate current.

The amplification factor is a dimensionless number. It ranges from several units to several tens for various types of triodes.

Sometimes, it is more convenient to use the quantity equal to the reciprocal of the amplification factor, known as the *penetration factor*:

$$D = 1/\mu = -\Delta V_g/V_p \quad \text{or} \quad D = 1/\mu = |\Delta V_g|/|\Delta V_p|$$
$$\text{at} \quad I_k = I_p = \text{const} \tag{14-4}$$

As is seen, the penetration factor shows how many times the grid voltage is more effective in controlling the cathode current than the plate voltage. For example, let a tube have $\mu = 10$ and $D = 1/\mu = 0.1$. This shows that the effect of the grid is ten times that of the plate.

The penetration factor characterizes the screening effect of the grid. In other words, it shows which part of the electric field set up by the plate voltage can pass through the grid in the direction of the cathode. A grid with closely spaced wires screens the cathode from the plate more effectively, so penetration factor is low and the amplification factor is high in this case. In contrast, in the case of

a grid with wires spaced farther apart, the penetration factor is high and the amplification factor is low.

The penetration factor also shows which part of the plate voltage should be added to the grid voltage for the triode to be replaced by an equivalent diode.

The electron flow in the grid-cathode space of the triode is affected by a resultant field due to the plate and grid voltages. In some cases, it is more convenient to replace the two electrodes of the triode by one equivalent electrode which would establish a field equal to the resultant field near the cathode. In this way, the triode is replaced by a diode whose plate current is equal to the plate current of the triode. The electrode substituting for the plate and grid, known as the *virtual anode*, is placed instead of the grid (Fig. 14-10).

The voltage which should be applied across the electrodes of the equivalent diode to produce the same plate current, I_p, is termed the *composite controlling voltage*, V_d. At the cathode, this voltage should establish an electric field of the same strength as that produced in the triode by V_p and V_g taken together. Multiplying the plate voltage by the penetration factor, $V_p \times D$, we obtain that part of the plate voltage of the triode which, if applied across the electrodes of the equivalent diode, would produce the same component of the electric field near the cathode as the plate voltage in the triode.

Accordingly, the composite controlling voltage of the equivalent diode is

$$V_d = V_g + DV_p \tag{14-5}$$

The product of g_m, R_p and D is

$$g_m R_p D = (\Delta I_p/\Delta V_g)\,(\Delta V_p/\Delta I_p)\,(\Delta V_g/\Delta V_p) = 1 \tag{14-6}$$

This equation relating the three tube constants is called the *parameter equation of a triode.*

Substituting $1/\mu$ for D in the last equation, we obtain the parameter equation of a triode in another form:

$$\mu = g_m R_p \tag{14-7}$$

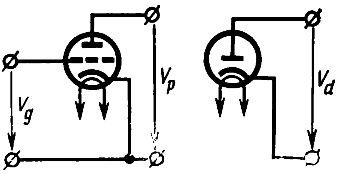


Fig. 14-10. Vacuum triode and an equivalent diode

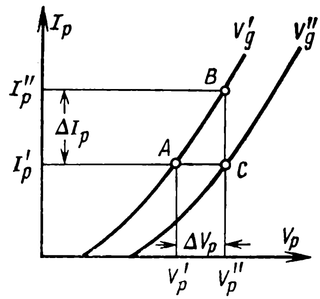


Fig. 14-11. Determining the parameters of a vacuum triode from its plate characteristics

As is seen, Eq. (14-7) makes it possible to determine any of the three quantities, provided that the other two are known.

The main parameters of a triode for specified values of V_p', I_p' and V_g' can be determined graphically from a family of plate or grid-plate characteristics obtained experimentally by the circuit shown in Fig. 14-4 or taken from tube manuals.

To deduce the triode parameters from plate characteristics obtained at V_g' and V_g'' (Fig. 14-11), one uses the characteristic triangle *ABC* in which the apex *A* is located by the specified values of plate voltage and current, V_p' and I_p'.

The second apex, *C*, of the triangle is obtained at the point where the second plate characteristic is cut by the line drawn from the apex *A* parallel to the *x*-axis. The third apex, *B*, is located as the first plate characteristic is intersected by the line drawn from the apex *C* parallel to the *y*-axis. The sides of this right-angled triangle are formed by the increment in the plate current ΔI_p, and the increment in the plate voltage ΔV_p. An increment in the grid voltage is defined as the difference between the specified grid voltages $\Delta V_g = \Delta V_g'' - \Delta V_g'$, for which the characteristics are measured. Substituting the values of ΔI_p, ΔV_p and ΔV_g into Eqs. (14-1), (14-2), (14-3) or (14-4), we obtain g_m, R_p, and μ or D of the triode.

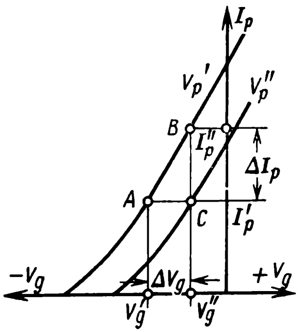


Fig. 14-12. Determining the parameters of a vacuum triode from its plate-grid characteristics

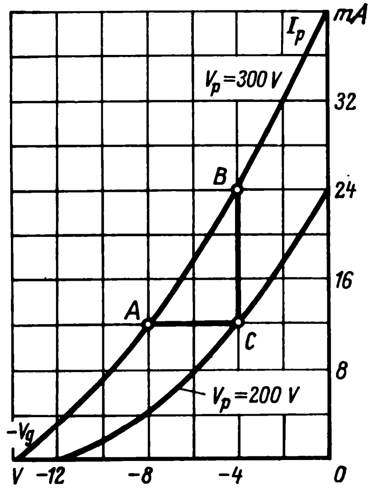


Fig. 14-13. To Example 14-1

If we use the grid-plate characteristics, the characteristic triangle *ABC* is constructed in a similar way (Fig. 14-12). The sides of this right-angled triangle are ΔV_g and ΔI_p. Here, an increment in the plate voltage is defined as the difference between the specified plate voltages $\Delta V_p = V_p' - V_p''$, at which the characteristics are measured.

Example 14-1. Find g_m, R_p and μ of a triode, using the characteristic triangle *ABC* (Fig. 14-13).

Solution.

1. The mutual conductance is

$$g_m = \Delta I_p/\Delta V_g = (24 - 12)/(8 - 4) = 12/4 = 3 \text{ mA/V}$$

2. The dynamic plate resistance is

$$R_p = \Delta V_p/\Delta I_p = (300 - 200)/12 = 100/12$$
$$= 8.33 \text{ kilohms}$$

3. The amplification factor is

$$\mu = \Delta V_p/\Delta V_g = (300 - 200)/(8 - 4) = 100/4 = 25$$

4. The penetration factor is

$$D = 1/\mu = 1/25 = 0.04$$

14-4. Interelectrode Capacitance of the Vacuum Triode

A vacuum triode has three metal electrodes spaced very small distances apart. Because of this, each pair acts as a small capacitance, called an *interelectrode capacitance*.

The vacuum triode has three interelectrode capacitances, namely, the grid-to-cathode (or input) capacitance, C_{gk}, the plate-to-cathode (or output) capacitance, C_{pk}, and the plate-to-grid (or transfer) capacitance, C_{pg} (Fig. 14-14).

These interelectrode capacitances are no fixed quantities, but vary, depending on the size and form of the electrodes, the spacing between them, etc. In low- and medium-power tubes they range from a few to twenty picofarads; in high-power tubes they may run up to fifty picofarads.

The capacitive susceptance, $\omega C = 2\pi fC$, is proportional to frequency, so an increase in frequency is accompanied by an increase in capacitive currents which may reach high values, thereby impairing the performance of a triode as an amplifier.

For example, C_{gk} causes the capacitive current in the grid-cathode-signal source circuit. The capacitive current gives rise to a voltage drop across the internal resistance

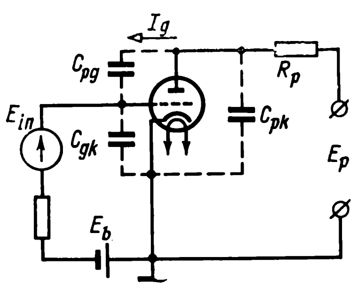


Fig. 14-14. Interelectrode capacitances of a vacuum triode

of the signal source, R_s, and reduces the input voltage across the triode. A consequence of this is a reduction in the output voltage and power output. The negative effect of C_{pg} is that its reactance, $1/\omega C$, goes down with increasing frequency, and the a.c. (signal) voltage gives rise to a current flowing from the plate circuit through this capacitance into the grid circuit (Fig. 14-14). Thus, feedback takes place, which adversely affects the performance of the triode.

14-5. Types of Vacuum Triodes

There exist two basic types of vacuum triodes. They are power-amplifying (modulator) triodes designed to amplify voltage and power, and transmitting triodes intended to generate electric oscillations.

The amplification factor of transmitting triodes, μ, ranges from unities to tens, their dynamic (a.c.) resistance, R_p, from a few to tens of kilohms, and their mutual conductance, g_m, from units to several tens of milliamperes per volt.

Physically, vacuum triodes can be divided into single and double triodes. A double triode is a tube which incorporates two triodes in the same envelope.

Soviet-made triodes are classed according to USSR State Standard GOST 13393-67 (see Sec. 13-6), the letter "*C*" standing for a single triode, and the letter "*H*" for a double triode. The envelope and tube sizes are classed in the same way as in Sec. 13-6, *b*.

Figures 14-15 and 14-16 represent the symbols for double triodes.

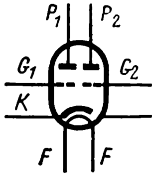


Fig. 14-15. Common-cathode double triode

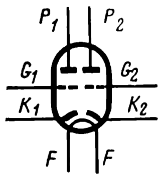


Fig. 14-16. Split-cathode double triode

The main advantages of vacuum triodes are that they are simple in design and reliable in operation. Besides, their grid-plate characteristic has a large linear portion.

Among the main drawbacks of vacuum triodes are the low amplification factor, μ, and high grid-to-plate capacitance, C_{pg}, due to which fact they are seldom used at high frequencies.

14-6. Vacuum Tetrodes

(a) Principle of Operation

The amplification factor of a vacuum triode cannot be increased by making the grid denser, because, if so, the cut-off voltage would go down and the grid-plate characteristics of the triode would become positive. Then, the tube would not operate as an amplifier, as it would draw grid currents.

However, this can be avoided by placing a second, or, screen, grid between the plate and the first (or control) grid.

Such tubes having both the screen and control grids are known as four-electrode tubes, or *tetrodes* (Fig. 14-17).

A tetrode has a control grid with the wires spaced far apart, due to which fact the negative cut-off voltage of the tube is high and, therefore, the tube has negative grid-plate characteristics. The wires of the screen grid are closely spaced, so it shields the cathode and control grid from the plate field and neutralizes the plate field near the cath-

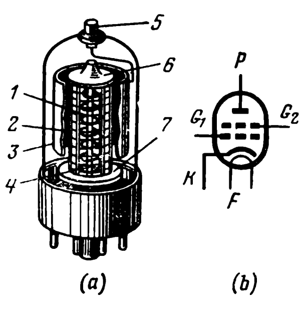


Fig. 14-17. Vacuum tetrode
(*a*) construction; (*b*) diagram symbol; *1*—control grid; *2*—screen grid; *3*—plate (anode); *4*—cathode; *5*—plate (anode) cap; *6*—upper shield; *7*—lower shield

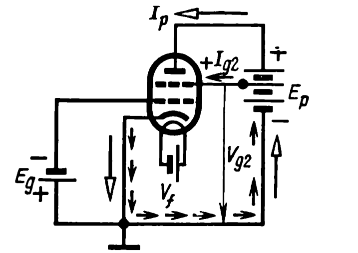


Fig. 14-18. Connection of a vacuum tetrode in a circuit

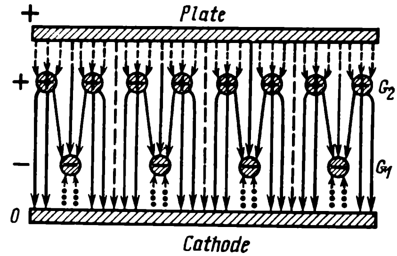


Fig. 14-19. Simplified map of the electric field in a vacuum tetrode

ode. The effect of the plate voltage on the electric field strength near the cathode is also small, because this field is produced by the screen grid, for which purpose a positive voltage of $V_{g2} \leqslant 0.5\ V_p$ is applied to it (Fig. 14-18).

Some electrons pass through the screen grid, reach the plate and form the plate current. Other electrons strike the screen grid and produce the grid current, I_{g2}, which must be as low as possible.

A map of the electric field in a tetrode is shown in Fig. 14-19. For simplicity, it does not show the fields due to the two grids. As the screen is a denser mesh than the control grid and has a lower potential than the plate, it intercepts most of the electric lines of force due to the plate. A small part of the original plate field reaches the control grid, and a still smaller part reaches the cathode.

Owing to the decrease in the field strength between the plate and control grid, the grid-to-plate capacitance C_{pg} between these electrodes is reduced to a few tenths or hundredths.

A fall in the plate field near the cathode reduces the effect of the plate voltage on the plate current, while the effect of the control grid potential on the plate current remains the same, for there is no screen between the control grid and the cathode. Therefore, the amplification factor μ and dynamic plate resistance R_p of the tetrode are considerably higher than those of the triode, while the mutual conductance g_m is the same.

(b) The Dynatron Effect

The plate characteristic of a tetrode is the relation between plate current and plate voltage with grid voltages held constant (Fig. 14-20), that is, $I_p = f(V_p)$, with V_{g1} and V_{g2} being constant.

The grid-plate characteristic of a tetrode is the relation between screen-grid current I_{g2} and plate voltage with grid voltages being constant (Fig. 14-20), that is, $I_{g2} = f(V_p)$, with V_{g1} and V_{g2} held constant.

Let us apply normal voltages V_{g1} and V_{g2} to the grids of a tetrode and gradually increase the plate voltage from zero.

At the zero plate voltage, all electrons passing through the control grid reach the screen grid and generate a grid current, I_{g2}. This takes place because the screen grid has a positive potential while plate current is zero due to the zero potential at the plate.

When the plate voltage increases to about 20 V, the plate current rises too, while the screen-grid current goes down (Fig. 14-20, region *I*). In this case, the plate characteristic rises and the grid-plate characteristic falls.

Further increase in the plate voltage results in a rise in electron energy and secondary emission. Secondary electrons move back to the screen grid whose potential is higher than that of the plate, so the plate current goes down and the grid current I_{g2} rises (Fig. 14-20, region *II*). This is called the *dynatron effect*. This effect reaches its maximum at the boundary between regions *II* and *III* (Fig. 14-20).

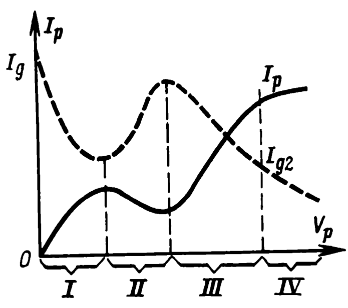


Fig. 14-20. Plate and grid-plate [$I_{g2} = f(V_p)$] characteristics of a vacuum tetrode

The dynatron effect ceases when the plate voltage becomes equal to the screen voltage, $V_p = V_{g2}$; this takes place at the boundary of regions *III* and *IV*. When the plate voltage exceeds the grid voltage, secondary electrons begin to return to the plate and the plate current rises again with increasing plate voltage, while the grid current grows smaller (Fig. 14-20, region *III*).

The dynatron effect makes it difficult to use tetrodes in amplifying circuits as it causes instability of operation and high distortion.

(c) The Beam-Tetrode

In a *beam-power tetrode* (or simply, a beam-tetrode), the dynatron effect is eliminated by a negative space charge between the plate P and the screen grid *2*, the field of which

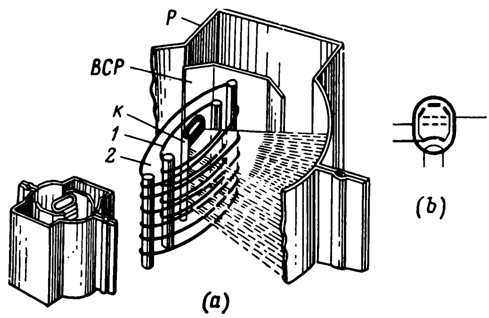


Fig. 14-21. Beam-power tetrode
(*a*) construction and (*b*) diagram symbol

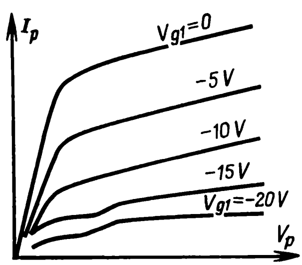


Fig. 14-22. Plate characteristics of a beam-power tetrode

decelarates electrons and drives them back to the plate (Fig. 14-21). For this purpose, the cathode K is made flat, the wires of the control (*1*) and screen (*2*) grids have the same spacings, and they are levelled up. The beam-tetrode has two more parts—metal beam-confining plates or shields, BCP, which are located between the screen grid, *2*, and plate, P, and connected to the cathode.

In the beam-tetrode, the electrons travelling from the cathode to the plate pass between the grid wires in a series of fan-shaped beams (sheets or layers). The beam-confining plates make the electrons concentrate on the cylindrical parts of the plate. Due to this concentration of electron beams, a negative space charge eliminating the dynatron effect is formed.

Figure 14-22 shows the plate characteristics of a beam-tetrode, which illustrate that the dynatron effect in it is considerably smaller.

The amplification factor of a beam-tetrode is about several hundred, its dynamic plate resistance ranges from tens to hundreds of kilohms, its mutual conductance varies from a few to several tens of milliamperes per volt, and its grid-to-plate capacitance C_{pg_1} is approximately equal to 0.1 pF or 0.2 pF.

The advantage of the beam-tetrode is the low screen-grid current which does not exceed 10 per cent of plate current.

14-7. Pentodes

The pentode is a five-electrode tube having a plate, a cathode and three grids: a control grid G_1, a screen grid G_2 and a suppressor grid G_3 (Fig. 14-23).

The suppressor grid G_3 located between the screen grid and plate is connected to the cathode and, therefore, is at zero potential, $\varphi_{g3} = 0$. The potentials at points in the field near the suppressor grid are lower than the plate potential, so the secondary electrons moving away from the plate run into a retarding field and return to the plate. In this way, the dynatron effect is eliminated.

In pentodes, the wires of the screen grid are closely spaced and those of the suppressor grid are spaced wider apart than the control grid in accordance with their purpose.

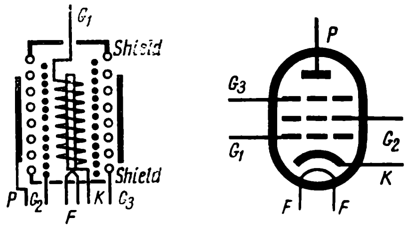


Fig. 14-23. Construction of an r.f. pentode and its diagram symbol

The screen grid is intended to reduce the grid-to-plate capacitance, and the suppressor grid is used to bring down slightly the potentials at the points near it, which ensures the return of the secondary electrons to the emitting surface and, at the same time, does not offer considerable resistance to the primary electrons emitted by the cathode.

The plate characteristic of the pentode has no dips due to the dynatron effect and is similar to that of the beam-tetrode (Fig. 14-22). At first, the plate current rises sharply with increasing plate voltage due to redistribution of currents between the plate and screen grid. At a higher plate voltage, the current rises slowly because the third grid in the pentode reduces the effect of the plate voltage on the tube field and plate current still more. That is why the amplification factor and dynamic plate resistance of pentodes reach such high values, the former being as high as one thousand and more and the latter being 1 or 2 megohms. The mutual conductance is low as compared with the triode, ranging from a few to tens of milliamperes per volt.

Due to the very dense screen and the additional suppressor grid, the grid-plate capacitance of r.f. pentodes does not exceed 0.003 or 0.004 pF, which is far less in comparison with the tetrode. Such pentodes are widely used at radio frequencies and are the main receiving-amplifying and transmitting tubes.

The r.f. pentodes whose grid-plate characteristics have two portions—a long gradual portion, *1*, and a steep rising portion, *2* (Fig. 14-24)—are known as *variable-mu* or *remote cut-off tubes*. This type of characteristic is obtained by wind-

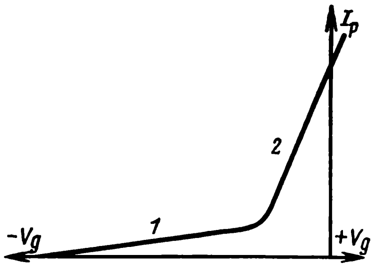


Fig. 14-24. Plate-grid characteristic of a variable-μ tube

ing the control-grid wires with variable spacing, usually a wide spacing in the centre (a "window") and a close spacing at the ends. Accordingly, the penetration factor of different parts of the grid is not the same. At a very negative grid voltage, the tube is cut off by the closely spaced segments of the grid, but remains conducting through the "window" which has a low amplification factor and insignificant conductance (Fig. 14-24, portion *1*). At a low negative grid voltage, the marginal areas of the grid having a high amplification factor and mutual conductance are rendered conducting (portion *2*).

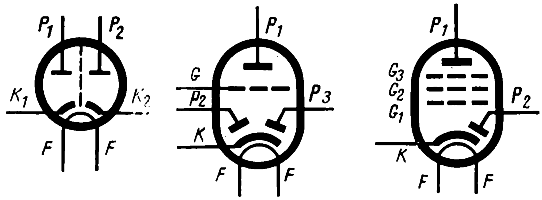
As is seen, the mutual conductance of these pentodes depends on the grid voltage. Such tubes are widely used, for example, in the automatic gain control circuits of radio receivers.

14-8. Multiple-Unit and Multi-Grid Tubes

There exist a great variety of *multiple-unit* and *multi-grid tubes* with which one can reduce the size and simplify the circuitry of electronic equipment.

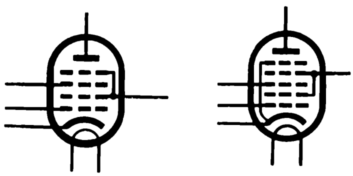
A multiple-unit tube contains several groups of electrodes associated with independent electron streams within one envelope. Examples of such tubes are shown in Fig. 14-25. Use is often made of double diodes (*a*), double diode-triodes (*b*), diode-pentodes (*c*), double triodes with a common cathode (Fig. 14-15), and double triodes with separate cathodes (Fig. 14-16).

A multi-grid tube contains a plate, a cathode and four or more grids. In these tubes, several electrodes can carry



(a) (b) (c)

Fig. 14-25. Diagram symbols of multiple-unit tubes
(*a*) double diode; (*b*) double diode-triode; (*c*) diode-pentode



(a) (b)

Fig. 14-26. Diagram symbols of multi-grid tubes
(*a*) hexode; (*b*) heptode

out the same function. For example, the plate current can be controlled by two voltages applied to separate grids. The influence exerted by one control grid upon the other is reduced by an additional grid placed between them. Tubes with six electrodes are called *hexodes* (Fig. 14-26,*a*), those having seven electrodes are known as *heptodes* (Fig. 14-26,*b*), and eight-electrode tubes are termed *octodes*.

As has been stated in Secs 13-6 and 14-5, according to USSR State Standard GOST 13393-67, the second element in a tube designation indicates the type of the tube. In designations for Soviet-made tetrodes, pentodes and multiple-unit tubes, П stands for an output pentode or a beam-power tetrode; К, for an r.f. variable-mu pentode; Ж for an r.f. pentode; А, for a frequency-converter tube with two control grids, apart from a pentode; Г, for a diode-triode; Б, for a diode-pentode; Ф, for a triode-pentode; И, for a triode-hexode, a triode-heptode and a triode-octode; Р, for a double tetrode or a pentode (Russian letters throughout).

Chapter Fifteen Gas-Filled Tubes

15-1. Electric Discharges in Gases and Their Volt-Ampere Characteristics

As their name implies, all gas-filled tubes have an amount of gas (or mercury vapour). When the cathode of a gas-filled tube is heated, it emits a very large supply of electrons which encounter the atoms of the gas or mercury vapour. As this takes place, the electrons give up some part of their energy ($mv^2/2$) to the gas atoms. If the velocity of these electrons is below some definite level, their encounter will be elastic. Otherwise, this collision will be inelastic, and the energy obtained by the atoms will cause their excitation or ionization. In an excited atom, one or more electrons are in unstable energy states (or levels) that are higher than their normal energies. As this happens, the electrons emit radiation (a quantum of light) which causes the gas to glow. The energy necessary to excite an atom depends on the excitation potential, V_{exc}

$$mv_{\max}^2/2 \geqslant W_{exc} = V_{exc}e \tag{15-1}$$

For example, the excitation potential for helium is 20.8 V and that for mercury vapour is 4.9 V.

The energy necessary for ionization, that is, for splitting an atom into an ion and electrons, depends on the ionization potential, V_{ion}, which is higher than the excitation potential, namely

$$mv_{\max}^2/2 \geqslant W_{ion} = V_{ion}e > V_{exc}e \tag{15-2}$$

For example, the ionization potential for helium is 24.5 V and that for mercury vapour 10.4 V.

Under normal conditions, any gas contains an insignificant number of electrons and ions (charge carriers), so gases are good dielectrics.

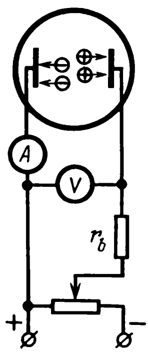


Fig. 15-1. Test set-up to measure the volt-ampere characteristic of a discharge gap

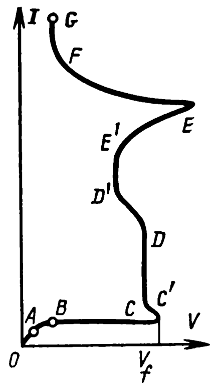


Fig. 15-2. Volt-ampere characteristic of a discharge gap

A gas can be made conductive through ionization by radioactive or cosmic rays, by an electric field, or high temperature. When a gas is continuously ionized at a constant rate, the splitting of atoms into electrons and ions is accompanied by partial recombination, so that the number of electrons per unit volume remains constant.

If we apply a voltage to the electrodes of a gas-filled tube as shown in Fig. 15-1, the electric field will cause the charge carriers of the ionized gas, that is, positive ions and electrons, to move. The ions will move with, and the electrons against the field. As the applied voltage rises, they will flow more rapidly.

The operation of gas-filled tubes is based on the *electric discharge* caused by the flow of an electric current through the filling gas.

At first, the current increases in proportion to the voltage across the electrodes (Fig. 15-2, section *OA* on the volt-ampere characteristic). Then, the rate of current rise slows down, and finally the current ceases growing (sections *AB* and *BC*). This means that all ions are transferred from one electrode to the other without recombination. This current is known as *saturation current*. If we increase the voltage further, the current will start rising again, at first slowly,

then, at what is known as the *firing voltage*, V_f, suddenly. At this voltage, the speed of the electrons and their kinetic energy due to the increased field strength are enough to ionize neutral atoms by collision. So, *impact* or *collision* ionization takes place. In turn the secondary charges accelerated by the field produce further ionization, so that the number of electrons and ions rises cumulatively. The space between the electrodes is filled by an ionized gas, or the gas plasma, which has high conductivity. Due to this fact, the current rapidly rises and the voltage falls (Fig. 15-2, section $C'D'$ on the volt-ampere characteristic). When the number of free electrons quickly rises, a *glow discharge* takes place, revealed by a slight luminosity and hissing of the gas. Now the discharge becomes a self-maintaining one, because, no external ionizer is required to maintain it. The current density rises to about 10^{-3} A/cm^2.

The most commonly used gas-filled devices utilizing glow discharge are neon-lamps, *VR* (voltage-regulator or voltage-reference) tubes, also known as stabilizer diodes, and thyratrons.

Glow discharge is preceded by the *Townsend* or *dark discharge* at which the current density is considerably less, being about 10^{-6} A/cm^2.

An *arc discharge* occurs at a current density much greater than a glow-discharge—up to 10^2 A/cm^2, and at a voltage of about 15 to 30 V. Arc discharges may occur not only in a rarefied gas but also under normal atmospheric pressure.

Arc discharges can be initiated in different ways.

(a) When the voltage in the interelectrode space reaches a certain value, called the *arc firing voltage*, V_{af} (Fig. 15-2, point E), the glow discharge turns into an arc (Fig. 15-2, portion FG). In this case, the arc is maintained by thermionic emission from the cathode heated by ion bombardment. This is known as a *self-maintaining discharge*. Owing to thermionic emission, the number of electrons increases, and so does the arc current. As a result, the value of voltage drop across the ballast resistor grows and the voltage across the electrodes goes down.

If the cathode is heated by a current from an external source, this process is called a *nonself-maintaining arc discharge*.

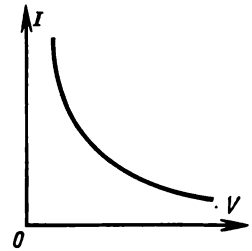


Fig. 15-3. Volt-ampere characteristic of an arc discharge

(b) An arc discharge can be initiated by bringing the electrodes together until they touch each other. The electrodes are strongly heated at the point of contact, so the interelectrode space is ionized and an arc strikes when the electrodes are separated. The gas plasma between the electrodes has a very high temperature (over 4000°C) and high conductivity. As the current increases, the temperature and conductivity of the plasma rise but the voltage across the electrodes goes down, thus the arc has a drooping volt-ampere characteristic (Fig. 15-3).

In addition to the arc discharge due to thermionic emission from the cathode, the arc discharge can be caused by field emission from a liquid mercury, or mercury-pool, cathode (such as in mercury-vapour tubes). In this case, the arc discharge takes place in mercury vapour. The base of the arc is a luminous spot on the surface of the mercury pool which emits electrons ionizing the mercury vapour. The phenomenon of an electric arc was discovered by the Russian academician V. V. Petrov in 1802.

In addition to the three main types of gas discharge considered above, two more varieties can be distinguished.

Corona Discharge. It is an easy matter to obtain a very strong electric field on the surface of small-diameter wires or wire points. When the field strength reaches a certain value, the Townsend discharge occurs and a weak glow known as a *corona* is observed. The corona discharge is initiated by gas ionization.

Spark Discharge. If the voltage across and the field strength in the air gap between the electrodes reach the breakdown values (see Sec. 1-7), a spark discharge takes

place. It appears as a brightly glowing tortuous channel connecting the electrodes. In this channel, the avalanche of electrons and ions follows the line of least resistance causing the temperature and pressure to rise sharply, so the spark discharge is accompanied by crackling sound.

15-2. Nonself-Maintaining Arc-Discharge Devices

(a) Gas-Filled Thermionic Diodes

A gas-filled thermionic diode or, simply, a *gas diode* is intended to rectify alternating current. It uses as a filler a small amount of mercury vapour or inert gas introduced under a pressure of 15 to 70 Pa (0.1 to 0.5 mm Hg) after its glass or metal envelope has been evacuated. The gas diode (Fig. 15-4) has two electrodes: nickel or graphite anode and an oxide-coated tungsten cathode. In power gas diodes, the cathode is put inside a metal container (a heat-conserving cathode, as it is called). The container greatly decreases the radiation of the heat, thereby saving filament power.

The cathode is supplied by a filament transformer. The filament voltage should not be higher than 5 V because, if the voltage exceeds this level and the ionization potential is low (about 10 V for mercury vapour), an arc may strike between the cathode ends. Therefore, the filament current should be high, from unities to a few tens of amperes, so it takes more time (from several minutes to a few tens) to warm up the cathode.

At first, when the anode voltage rises from zero, a purely electronic current flows in the gas diode because the velocity of electrons in the weak electric field is insufficient for them to ionize the gas. This process is represented by portion *OA* in the volt-ampere characteristic of the gas diode (Fig. 15-5). If the anode voltage is a little higher than the ionization potential, the electrons escaping from the cathode will be accelerated by the electric field to a velocity sufficient to excite and ionize the gas (or mercury vapour).

The ionized gas turns into a plasma, and an arc discharge takes place. This instant is represented by point *A* on the characteristic of the gas diode.

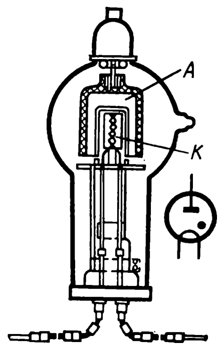


Fig. 15-4. Gas diode and its diagram symbol

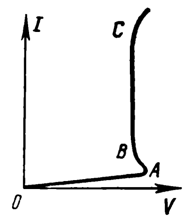


Fig. 15-5. Volt-ampere characteristic of a gas diode

When the positive ions of the gas neutralize the negative space charge near the cathode, the electron emission slightly rises.

The current rise caused by a decrease in the resistance of the anode circuit due to a load change, or an increased supply voltage, has almost no effect on the voltage drop between the anode and cathode.

The operation of a gas diode under load is represented by portion *BC* of its volt-ampere characteristic. The voltage and current ought not to be increased further (the limit is shown by point *B*), because it may damage the tube.

The advantage of a gas over a vacuum diode is a lower voltage drop, so rectifiers using gas diodes have a greater efficiency.

As has already been said the gas diode needs a long warm-up period before the anode voltage may be applied, since otherwise the cathode may lose emission.

Figure 15-6 shows the voltage and current waveforms of a half-wave gas rectifier. At first, the anode grows more positive, until it reaches the firing (or breakdown) potential. At that instant, a discharge is established, and the anode potential suddenly decreases to a constant value. This condition is maintained until the end of the positive half-cycle, when the voltage drops to zero. The negative half-cycle of the anode voltage has the usual shape of a

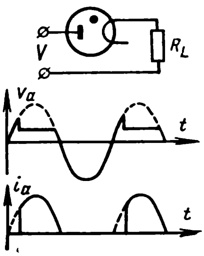


Fig. 15-6. Voltage and current waveforms of a gas diode used in a half-wave rectifier circuit

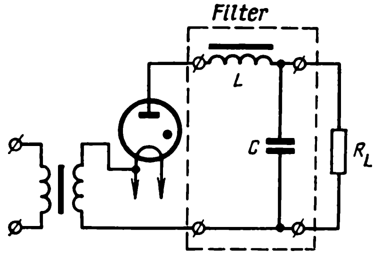


Fig. 15-7. Gas diode rectifier using a filter

half-sinusoid. The anode current appears as a train of sine-wave pulses with a clipped portion at the leading edge.

Gas diode rectifiers utilize the same circuits as those based on vacuum diodes, except that no capacitor should be connected at the input of the filter (Fig. 15-7), since otherwise the charge current of the capacitor may exceed the maximum safe value, $I_{a\text{-max}}$, and cause the cathode to lose emission, that is, damage the tube.

The parameters describing the performance of gas diodes are the filament voltage and current, the maximum and average rectified current, the voltage drop across the tube, the peak inverse voltage, and the warm-up period.

The maximum anode current should not exceed the cathode emission current.

The peak inverse voltage is the maximum negative anode voltage at which the gas diode is still capable of unidirectional conduction without an arc-back.

Gas-filled diodes are used in low-power rectifiers (such as for battery charging, to supply control circuit, etc.).

(b) Hot-Cathode Thyratrons

Like gas diodes, thyratrons are arc-discharge tubes, and have much in common with them. What sets thyratrons apart from gas diodes is that they are grid-controlled tubes, having a third electrode, the grid, controlling the instant at which the tube can fire or start conducting (Fig. 15-8). The cathode in the thyratron is enclosed by a baffle whose

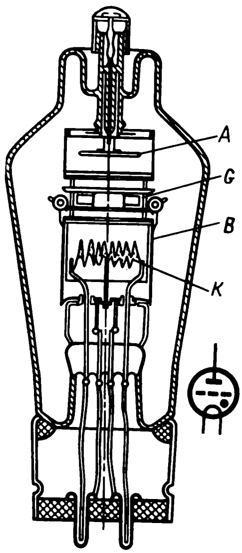


Fig. 15-8. Thyratron and its diagram symbol

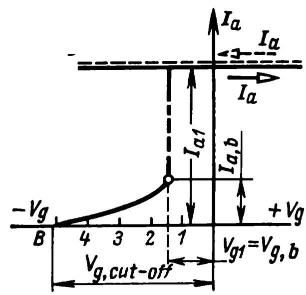


Fig. 15-9. Anode-grid characteristic of a hot-cathode thyratron

upper window is covered by a grid made in the shape of a perforated disc. The baffle does not allow an electric field to bypass the grid and establish itself in the anode-cathode space.

When a negative potential is applied to the grid, its electric field opposes the main field of the thyratron. At first, the applied negative potential is sufficient to keep the tube non-conducting, so that $V_g > V_{g,cut\text{-}off}$. Next, the grid is gradually made less negative until the grid voltage is somewhat lower than $V_{g,cut\text{-}off}$. In the circumstances, a very low current begins to flow in the anode circuit, rising with decreasing V_g, as in the vacuum triode (Fig. 15-9). When the grid voltage decreases to $V_g = V_{g,b}$, the speed of electrons becomes high enough for them to cause ionization, establish an arc discharge, and produce the plasma. Now, the anode current jumps up to I_{a1} (Fig. 15-9) whose value is determined by the load resistance, $R_a = R_L$, and anode

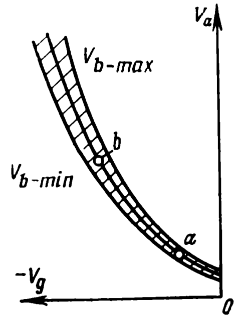


Fig. 15-10. Breakdown (grid-control starting) characteristic and range of a thyratron

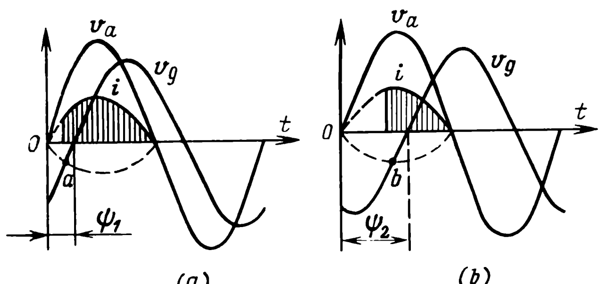


Fig. 15-11. Thyratron voltage and current waveforms for various phase differences between anode and grid voltages

voltage, V_a. In Fig. 15-9, the anode currents before and after the occurrence of a discharge are plotted to different scales.

When an arc discharge is established, the anode current no longer depends on the grid voltage. This is because, when the arc strikes, the grid collects a layer of positive ions which neutralize the negative charges on the grid, so it cannot control the anode current any longer. The arc can now be quenched only by reducing the anode voltage to about zero.

At a constant grid-cathode voltage, V_g, the arc strikes when the anode voltage, V_a, reaches what is known as the breakdown value. Hence, the anode voltage at which the

arc is established can be controlled by adjusting the grid voltage.

A plot of breakdown (or starting) voltage as a function of grid voltage (Fig. 15-10) is called the *breakdown* (or *grid-control starting*) *characteristic of a thyratron*. At a fixed grid voltage, a thyratron may fire at different anode voltages, ranging from $V_{b\text{-min}}$ to $V_{b\text{-max}}$. This is because the breakdown voltage depends on several variables, namely the pressure in the envelope, ambient temperature, grid circuit resistance, filament current, etc. So, instead of a single curve, it is usual to specify a range bounded by the curves for $V_{b\text{-min}}$ and $V_{b\text{-max}}$ (Fig. 15-10).

To hold the grid voltage at a safe value, a 1- to 100-ohm resistor is connected into the grid circuit.

When it operates in a rectifier circuit, a thyratron fires at a positive anode voltage and goes out at an anode voltage close to zero once during each cycle. Let us supply the grid with an a.c. voltage V_g, at the same frequency as the anode voltage V_a, but shifted in phase by an angle ψ_1 (Fig. 15-11,*a*). Firing will take place when the negative grid voltage has decreased and anode voltage has risen, so that both correspond to point *a* on the grid-control starting characteristic (Fig. 15-10). In Figure 15-11, the grid-control starting characteristic is represented by the dashed line.

By changing the phase of grid voltage, we can control the instant at which the thyratron fires (point *b* in Fig. 15-10). So, we can control the time that anode current can flow during each cycle, that is, the average current and voltage (Fig. 15-11,*b*).

Thyratrons are used in a.c. circuits operating at frequencies from 1 to 10 kHz. At higher frequencies, the tube would fail to recover in time and the grid would lose control.

They are employed in rectifiers, inverters, converters, automatic control systems, telemetry and telecontrol.

15-3. Glow-Discharge Devices

(a) Neon Indicator Lamps

The ability of gases to emit light on a glow discharge is utilized in discharge lamps. Those used in radio and electronic circuits are usually filled with *neon* or a *neon-argon*

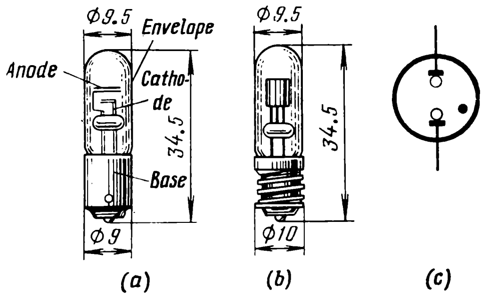


Fig. 15-12. Indicator lamps
(*a*) and (*b*) neon tubes and (c) their diagram symbol

mixture under a pressure of 2500 to 4000 Pa (20 to 30 mm Hg), and are mostly known as *neon indicator lamps*. They can be of the two-electrode and multi-electrode types. The latter are also called *numerical readout tubes* because their cathodes are given the shape of numerals.

The electrodes of a.c. two-electrode indicator lamps are made identical (such as discs); those of d.c. indicator lamps usually differ in shape.

Figure 15-12 shows two Soviet-made miniature neon lamps, types TH-0.2 and TH-0.3, intended for d.c. operation. Their power rating is fractions of a watt.

In the TH-0.2 indicator lamp (Fig. 15-12,*a*), the anode is a ring and the cathode is an oxide-coated disc.

In the TH-0.3 (Fig. 15-12,*b*), the anode is a wire 3 mm in diameter, and the cathode is an oxide-coated cylinder.

The breakdown voltage is always somewhat higher than the discharge voltage. So, to prevent indicator lamps from overload, ballast resistors are connected in series with them. Standard-size lamps (such as the Soviet-made TH-30) use ballast resistors built into the tube base; miniature devices use external ballast resistors.

Numerical readout tubes directly display the values of the quantity being measured as digits.

Figure 15-13,*a* shows a numerical readout tube containing ten tungsten-wire cathodes which are given the shape of the numerals from 0 to 9. The anode is made up of a fine-wire mesh placed in front of all cathodes (Fig. 15-13,*b*).

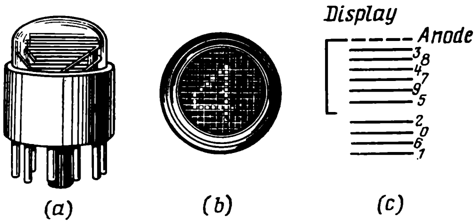


Fig. 15-13. Numerical readout tube
(a) external appearance; (b) end view; (c) arrangement of electrodes

It is connected to the side screen surrounding the cathodes, so that the distance between the anode and each of the cathodes is approximately the same.

The cathodes in the tube are arranged so that the "off" cathodes will not obstruct the view of those that are "on" (Fig. 15-13,*c*).

The cathodes are turned on by a switching circuit.

(b) VR Tubes

Glow discharge tubes intended to stabilize voltage across a load or in a d.c. circuit are known as VR-tubes or stabilizer diodes.

One type of VR tube has a glass envelope (Fig. 15-14) containing a cylindrical cathode K and a wire anode A arranged along the axis of the cathode. The envelope is filled with a mixture of inert gases (such as argon-neon or argon-helium) under a pressure of 2500 to 4000 Pa (20 to 30 mm Hg).

The cathode is made of steel, nickel or molybdenum and activated on the inside with barium or cesium to reduce its work function.

As is seen in Fig. 15-15, the normal glow region ABC on the volt-ampere characteristic of a VR tube runs almost parallel to the y-axis.

The VR tube is connected in parallel with a load resistor, R_L (Fig. 15-16), and the parallel circuit is connected in series with a ballast resistor R_b.

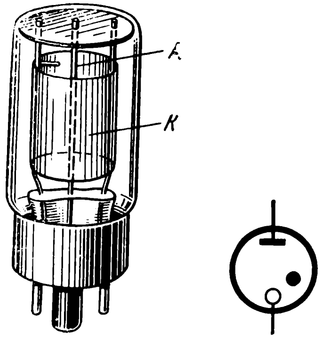


Fig. 15-14. VR tube and its diagram symbol

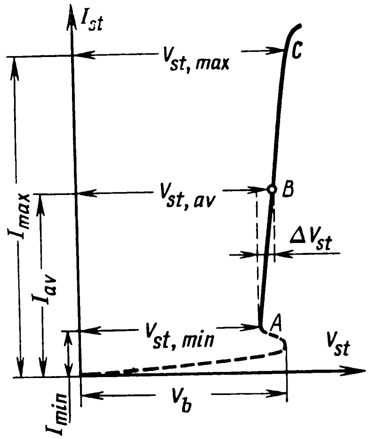


Fig. 15-15. Volt-ampere characteristic of a VR tube

The current in the ballast resistor is equal to that in the power source

$$I = I_{st} + I_L$$

The input voltage is the sum of the voltage drops across the ballast resistor, $V_b = IR_b$, and across the load resistor or VR tube

$$V_{in} = V_b + V_{st} = IR_b + V_{st}$$

Any change in the input voltage causes almost the same change in the voltage across the ballast resistor ($\Delta V_{in} \approx \approx \Delta V_b$), so an insignificant change can only occur in the voltage across the VR tube. This is because an incremental change in the VR tube voltage (ΔV_{st}) causes a considerable rise in the VR tube current (ΔI_{st}) and ballast resistor current (ΔI), and, consequently, a considerable rise in the voltage across the ballast resistor (ΔV_b).

When the input voltage is held constant, an increase in the load current results in a decrease in the VR tube current and an insignificant change in the load voltage.

The performance of a VR tube is characterized in terms of the stabilization factor which shows how many times

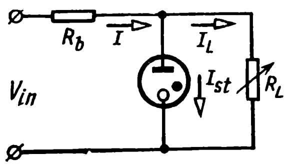


Fig. 15-16. Circuit of a stabilizer using a VR tube

a relative change in input voltage is greater than a relative change in load voltage:

$$k_{st} = \frac{\Delta V_{in}/V_{in}}{\Delta V_L/V_L} = \Delta V_{in} V_L / \Delta V_L V_{in} \qquad (15\text{-}3)$$

Where high stabilized voltages are required, use is made of several VR tubes connected in series.

VR tubes are manufactured for 70 V and higher and for currents ranging from 5 to 40 mA.

(c) Barretters

The *barretter* is a tube intended to maintain a constant current over a given supply voltage variation.

A barretter consists of a glass envelope containing an iron- or tungsten-wire filament. The envelope is filled with hydrogen under a pressure of 6000 to 25,000 Pa (50 to 200 mm Hg).

The cooling and heating conditions of the barretter filament are chosen such that any variation in the filament voltage produces almost a proportional change in its resistance. Thus, within certain limits, voltage variations result in insignificant current changes.

A barretter connected in series with a load (Fig. 15-17) will maintain the load current at an almost constant value despite considerable variations in supply voltage. If the

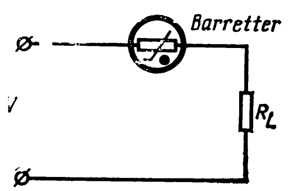


Fig. 15-17. Connection of a barretter in a circuit

load resistance is constant, both the current through and voltage across the load will be maintained almost constant over supply voltage variations.

Barretters can be used in both d.c. and a.c. circuits, because they do not respond to sudden variations in current, as they have a considerable time lag (ranging from 1 to 3 min).

Barretters have found wide application in long-distance communication circuits where they are used to stabilize the filament current of vacuum tubes.

(d) Glow or Cold-Cathode Thyratrons

Glow thyratrons are also known as *cold-cathode thyratrons.* The simplest cold-cathode thyratron consists of an envelope containing three electrodes (Fig. 15-18), namely the anode A, cathode K, and control grid (or trigger electrode) G. The filling gas and its pressure are the same as in the VR tube.

Its atomic-film cathode is made in the form of a cylinder. The plate is made of molybdenum and is given the shape of a rod. The nickel grid is a ring or a hollow cylinder enclosing the plate.

The power source E_a (Fig. 15-19) sets up a grid voltage, V_g, across the gap between the cathode and grid, so that the starting Townsend discharge producing a low initial ionization takes place in this space. When a positive pulse is applied to the circuit, the grid current I_g (Fig. 15-20, segment AB) rises by ΔI_g (segment BC), which causes the Townsend discharge to turn into a glow discharge and reach the anode if the anode-to-cathode voltage V_a is high enough to maintain the discharge.

After the thyratron has fired, the grid loses all control over the thyratron current.

The grid current I_g ranges from a few to tens of microamperes, while the anode current I_a varies from a few to tens of milliamperes.

The relation between the breakdown voltage V_b and grid current I_g is shown by the breakdown characteristic of a glow thyratron (Fig. 15-20). As I_g is increased, the level of ionization in the cathode-anode space rises and V_b goes down. However, V_b cannot be lower than V_m, the maintain-

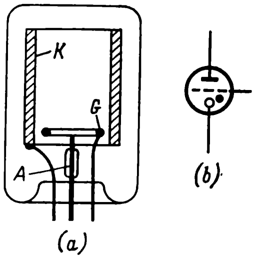


Fig. 15-18. (*a*) Glow-discharge (cold-cathode) thyratron and (*b*) its diagram symbol

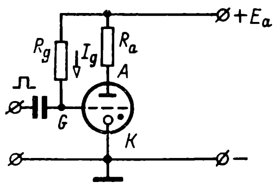


Fig. 15-19. Connection of a cold-cathode thyratron in a circuit

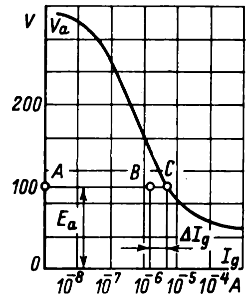


Fig. 15-20. Breakdown characteristic of a cold-cathode thyratron

ing voltage, that is, the one which maintains the glow discharge across the main gap.

The glow thyratron can be quenched by breaking the anode circuit or by reducing the anode voltage below the maintaining voltage.

The advantages of these thyratrons are their small size and mass, high mechanical strength, wide operating temperature range (from −60°C to +100°C), long service life, and low power consumption (they draw no filament voltage). The main drawback is the instability of their characteristics.

Glow thyratrons find use as relays in automatic control and other circuits.

15-4. Self-Maintaining Arc-Discharge Devices

(a) General

The main devices utilizing the self-maintaining (or self-sustaining) arc discharge are *mercury-arc rectifiers*, also known as *mercury-pool diode tubes.*

Mercury-arc rectifiers are the most commonly used gas-discharge power rectifiers.

Cathode emission is an important factor that limits the current-conducting capacity of thyratron tubes. The use of a mercury-pool cathode overcomes this limitation and extends the service life of a tube.

Modern mercury-arc rectifiers have metal envelopes.

According to the manner in which the starting arc is initiated and maintained at the cathode, mercury-arc rectifiers can be divided into *excitrons* and *ignitrons.* Apart from the main anode(s), an excitron has starting anodes which initiate an arc at starting and maintain it, especially when the load is disconnected. Ignitrons have no starting anodes, but they use a third electrode, the *igniter rod*, which produces a spark to establish a small localized arc before each positive half-cycle.

(b) The Excitron

At present, both multi-anode metal-tank rectifiers and sets of single-phase metal-tank rectifiers are manufactured designed for currents up to several thousands of amperes at medium voltages, and up to several hundreds of amperes at high voltages.

Figure 15-21 shows the schematic circuit and connection diagram of a three-anode three-phase excitron. The evacuated metal envelope carries a mercury-pool cathode K, three main anodes A_1, A_2 and A_3, and two auxiliary starting anodes A_s. The three main anodes are connected to the secondary windings of a three-phase transformer. These windings are star-connected and their neutral serves as the minus side of the load circuit. Power for the starting anodes A_s is supplied by an auxiliary transformer, Tr_{aux}. These anodes are used to maintain the starting arc irrespective of the load resistance R_L.

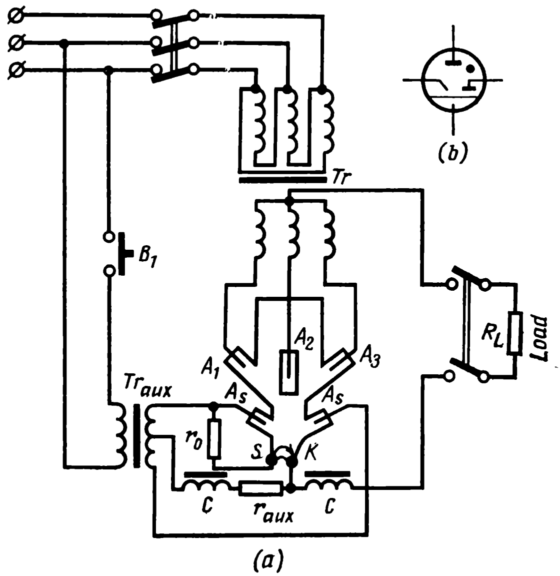


Fig. 15-21. (*a*) Construction and connection of a three-phase excitron and (*b*) its diagram symbol

The mercury rectifier is started by pressing the button B_1; voltage is thus applied to the auxiliary transformer. Due to the emf induced in the upper half of the secondary winding of this transformer, current will flow through resistor r_0, semiconductor igniter S, mercury-pool cathode K, resistor r_{aux}, and choke C. The semiconductor igniter, which is a rod made of boron carbide, is not wetted by mercury, that is why the current flow causes sparks between the rod and mercury pool, which initiate ionization. The electric field forces electrons to move from the cathode to the starting anode A_s, whose potential at this time is positive with respect to the cathode. While moving, the electrons ionize the mercury vapour which turns into a plasma, and an arc fires between the electrodes. When the potential at the first starting anode decreases and, at the same time, the potential at the second starting anode rises so that it exceeds that of the first one, the arc transfers from the first to the second starting anode.

The voltage drop in the plasma is small, so the most part of the voltage drop is across the layer between the mercury-pool cathode and the ion cloud which is formed at some distance from the cathode. The field strength in this layer reaches high values (about 10^6 V/cm), and the field causes the cathode to emit electrons. The source of this field emission is the glowing mercury spot constantly moving about on the surface of the cathode.

To maintain the cathode spot and, as a consequence, the arc, the anode current should not be less than 4 or 5 A.

When the voltage at one of the anodes falls below the critical value, choke C (Fig. 15-21) maintains the current required for the arc not to go out until a sufficient current starts flowing through another anode as the applied anode voltage alternates. Thus, a continuous cathode spot is maintained and current flows via the secondary winding of the auxiliary transformer, starting anode, cathode, resistor r_{aux}, and choke C to the neutral point of the secondary winding of the auxiliary transformer.

Each time the anode and cathode voltage changes sign, that is, during the existence of inverse voltage, there appears an insignificant inverse current between the electrodes.

When the three-phase transformer and load R_L are turned on, an arc is initiated between the cathode and main anode having the highest potential with respect to the cathode. Then, the arc is transferred to the second and third main anodes. Thus, current flows through each of the main anodes only for one-third of a cycle (see also Sec. 18-3).

The voltage drop across the tube during conduction is usually small (about 20 to 25 V).

Figure 15-22 shows the internal arrangement of a Soviet-made type PM-500 air-cooled metal-tank six-anode excitron intended for a medium current (500 A).

At the bottom, the tank, *1*, fitted with cooling fins, *2*, has a cathode bowl, *3*, holding the mercury pool, *4*. The cathode lead, *5*, igniter lead, *6*, and igniter, *7*, are located in the lower part of the bowl. The lid, *8*, gives support to six main anodes, *9*, with leads, *10*, and also to starting anodes, *11*, with leads, *12*. The deionization filter, *13*, limits ion flow to the plate, *9*, and control grid, *14* (with lead *15*) which controls the moment of firing. The screen,

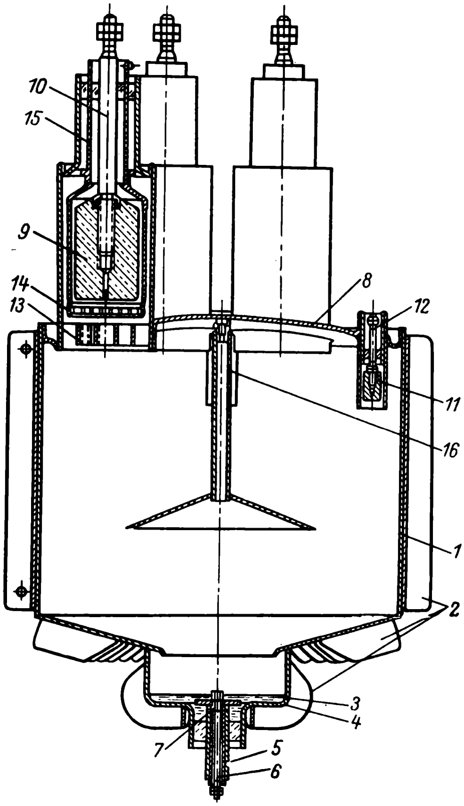


Fig. 15-22. Construction of an air-cooled metal-tank six-anode mercury-arc rectifier

16, protects the anodes and grids against exposure to the ascending mercury vapour jet.

Figure 15-23 is a sketch of the internal arrangement of a Soviet-made type PMHB-500 high-power single-anode excitron designed for a medium current (500 A).

The igniter, *1*, is a pointed rod made of boron carbide. The igniter conducts a short surge of current to the surface of the mercury pool and produces sparks which establish a small localized arc between the cathode and anode. The starting anode, *2*, is a graphite rod located in a separate sleeve, *4*. The graphite screen, *3*, protects the anode against exposure to the mercury jet or drops and, at the same time,

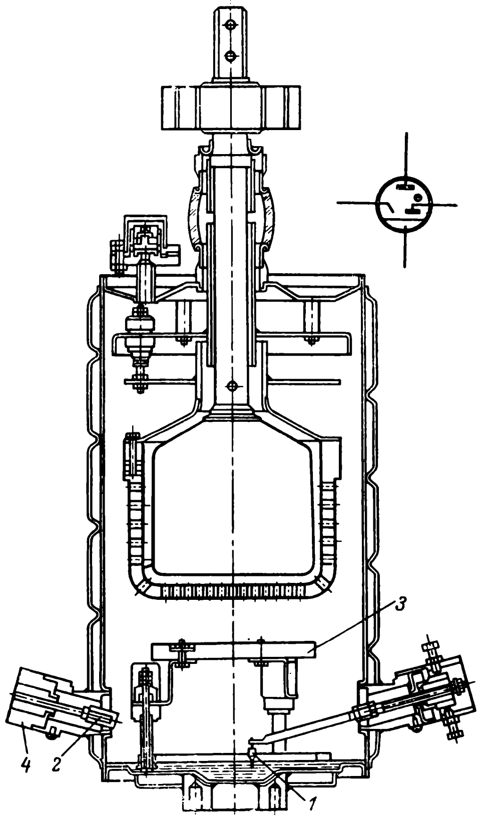


Fig. 15-23. Construction of a single-anode power excitron and its diagram symbol

serves as a deionization filter. The control grid encloses the anode from the bottom and sides. The tube utilizes water cooling applied through a spiral channel within the tank.

(c) The Ignitron

Figure 15-24 shows a low-power metal-tank ignitron and its diagram symbol. The tank, *1*, is a steel cylinder cooled by water circulating in a water jacket. The tank encloses a graphite anode, *2*, insulated by glass, *3*. The cathode is

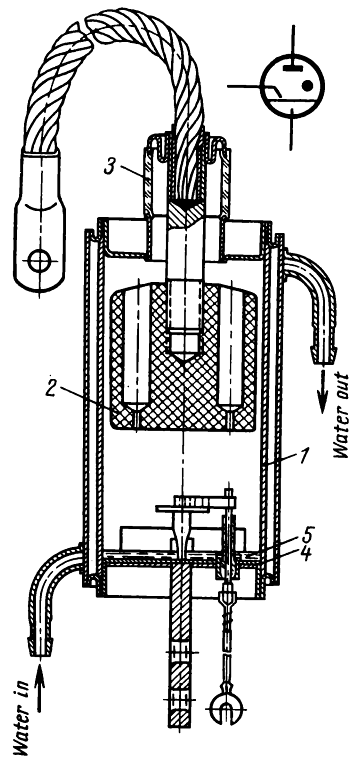


Fig. 15-24. Construction of a low-power, metal-tank ignitron and its diagram symbol

a metal bowl, *4*, containing the mercury pool, *5*. The igniter rod made of boron carbide starts an arc at the beginning of each anode voltage cycle.

The advantages of the ignitron are a low voltage drop (15 to 20 V) and a high efficiency (of up to 98 or 99%).

Ignitrons are used in rectifiers and welding equipment.

15-5. Nomenclature for Soviet-Made Gas-Discharge Tubes

In the Soviet Union, the type designation for a gas-discharge tube consists of three elements.

The first element is a letter which characterizes the type of the tube. For example, СГ stands for a VR tube; ТГ,

for a gas-filled thyratron; ТР, for a mercury-arc thyratron; ТГР, for a mixed-gas thyratron; ТХ, for a glow thyratron; ГГР, for a mixed-gas gas diode; ГХ, for a glow-discharge gas diode; ГГ, for a gas-filled gas diode; ГР, for a mercury-vapour gas diode; И, for an ignitron; Э, for an excitron; ИН, for a glow-discharge indicator tube; and СН, for a neon indicator tube.

The second element is a number which indicates the numerical sequence of the tube.

The third element differs in accordance with the type of the tube: (a) for low-power thyratrons and glow-discharge VR tubes, it is the letter which indicates the structure of the tube; (b) in the case of gas diodes, ignitrons and excitrons, it is a fraction, the numerator specifying the average current (in amperes), and the denominator the peak inverse voltage (in kilovolts).

For example, "ИН-1" is interpreted as a neon indicator tube, type one; "СГ-3С", as a VR tube, type three, in a glass envelope more than 22.5 mm in diameter; "Э1-10/1.5", as an excitron, type one (one anode, a metal envelope), 10-A average current, 1.5-kV peak inverse voltage; "ТГ1-0.1/0.3" as a gas-filled thyratron, type one, 0.1-A average current, 0.3-kV peak inverse voltage.

Chapter Sixteen Semiconductor Devices and Their Application

16-1. Intrinsic Conduction in Semiconductors

In terms of conductivity, semiconductors lie between conductors and insulators (or dielectrics). The resistivities of conductors, semiconductors and insulators are from 10^{-8} to 10^{-5} ohm m, 10^{-5} to 10^{7} ohm m and 10^{7} to 10^{16} ohm m, respectively.

The conductivity of semiconductors is strongly dependent on temperature, electric fields, incident light, applied pressure, etc. In contrast to conductors, they possess not only electron (or negative, N-type) conduction but also hole (positive or P-type) conduction.

At present, the most commonly used semiconductor materials are germanium, silicon, gallium arsenide, and selenium.

Under certain conditions atoms form a particular type of chemical linkage in which each atom contributes one electron to a shared pair in the same orbit (Fig. 16-1*a*). This type of chemical linkage is known as the covalent electron-pair bond. Graphically, it can be shown by two lines connecting the atoms (Fig. 16-1*b*). For example, the element germanium belongs to the fourth group of Mendeleev's periodic table of elements, so its atom has four electrons in the valence band. Consequently, each atom in a germanium crystal forms covalent bonds with four adjacent atoms (Fig. 16-1*c*).

At a temperature close to absolute zero, a germanium crystal free from impurities is a dielectric—all of its valence electrons are mutually linked, there are no free electrons, and the material cannot conduct an electric current. As the

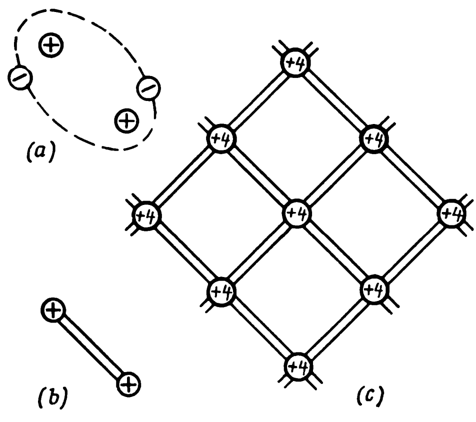


Fig. 16-1. Crystal lattice of a semiconductor
(*a*) covalent electron-pair bond; (*b*) graphical representation of a covalent bond; (*c*) bonds in the crystal lattice of germanium

temperature rises, or if the crystal is exposed to radiation, more valence electrons may acquire sufficient energy to break their covalent bonds and become free. Germanium behaves as a semiconductor already at room temperature. Under the action of an applied electric field, free electrons move and effect electron conduction.

Each missing electron constitutes in a covalent bond a vacant place known as a *hole.* Any of the electrons from a neighbouring bond may jump into the hole to restore the normal bond in that place, but it will leave a new hole behind which may be occupied by another electron, and so on. If a semiconductor is placed in an electric field, the field will cause the holes to move with the field, which is opposite to the direction in which electrons move. The motion of holes is equivalent to that of positive charges and constitutes a hole current. This is called hole conduction. In the case of electron conduction, a single free electron travels the entire path length in a crystal; in the case of hole conduction, a great number of electrons fill the vacancies in covalent bonds as if in a relay race, each travelling only a part of the total distance.

When covalent bonds in a crystal of pure semiconductor are broken, equal numbers of electrons and holes appear simultaneously in it.

Free electrons and holes appear in pairs, and the process is known as the generation of electron-hole pairs. Pair generation is inevitably accompanied by recombination of carrier pairs into neutral atoms. The generation and recombination of electron-hole pairs occur simultaneously, so there is a limit to the number of electron-hole pairs that can exist in a unit volume at a given temperature. For example, the number of electron-hole pairs (known as the density, concentration or population of charge carriers) per cubic centimetre of germanium at 20°C is $n =$ (approx.) 2.5×10^{13}, and the number of free electrons per cubic centimetre of a metal conductor is $n =$ (approx.) 10^{22}-10^{23}. As is seen, the conductivity of germanium at normal temperature is considerably lower than that of metals. As the temperature increases, the number of free electrons and holes rises, too, and the conductivity of germanium grows greatly.

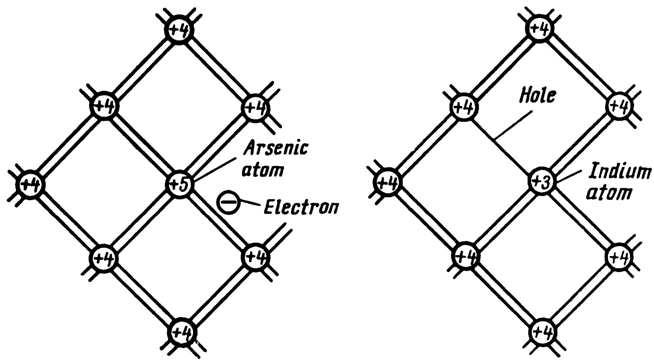
Semiconductors not containing any impurities are called *intrinsic*, or *I-type*, *semiconductors* and possess *intrinsic conduction*.

Semiconductor materials have a negative temperature coefficient of resistance, whose value is 10 to 20 times that of metals. When a metal is heated by 1°C, its resistance increases approximately by 0.4%, while in the case of a semiconductor, it falls by 4 to 8%. This property of semiconductors is utilized, for example, in thermistors whose resistance changes greatly at slight variations in temperature.

16-2. Impurity or Extrinsic Conduction in Semiconductors

The properties of a semiconductor can be changed by adding an insignificant amount of an impurity. If a semiconductor material is "doped" with atoms of some other element, it may come by an excess of free electrons over holes, or an excess of holes over free electrons.

For example, if one atom in the crystal lattice of germanium is replaced by an atom of arsenic which is a five-valence element, four electrons of the arsenic will form covalent bonds



(a) (b)

Fig. 16-2. Crystal lattice of an extrinsic semiconductor
(a) donor impurity (arsenic); (b) acceptor impurity (indium)

with three germanium atoms, while the fifth electron weakly connected to the arsenic atom becomes a free electron available for conduction (Fig. 16-2*a*). Thus, the arsenic impurity enhances electron conduction.

If indium, which is a three-valence element, is used to dope germanium, its three electrons will form covalent bonds with three germanium atoms. Now, there will be one incomplete bond on the fourth neighbouring germanium atom (Fig. 16-2*b*). All the four bonds will be completed if the missing electron is taken from a nearby germanium atom. But the electron jumping from that atom will leave behind a hole which may in turn be filled by an electron moving in from an adjacent germanium atom, and so on. What we have is an apparent flow of holes in the semiconductor. So germanium doped with indium has an increased hole conduction.

Semiconductors having predominantly electron conduction are N-type semiconductors (with "N" standing for "negative"). Semiconductors having predominantly hole conduction are called P-type (with "P" standing for positive). The charge carriers determining which type of conduction is predominant are called *majority carriers* (electrons in an N-type and holes in a P-type semicondutor). The im-

purities responsible for electron conduction are called *donor impurities* or simply donors, because they donate (or contribute) excess electrons to the host material. The impurities causing the prevalence of hole conduction, that is, those having fewer valence electrons than the host semiconductor, are called *acceptors*, because they accept excess electrons from the host. For germanium, examples of donors are arsenic, antimony and phosphorus, and examples of acceptors are indium, gallium and aluminium.

Any impurity radically affects the conductivity of the host material, so that it may increase ten to hundred thousand times, depending on the impurity concentration. For example, under normal conditions, one cubic centimetre of pure germanium contains about 4.2×10^{22} atoms and 2.5×10^{13} free electrons and holes. The addition of 0.001% of arsenic donates extra 10^{17} free electrons in the same volume, and the electron conductivity increases about ten thousand times.

16-3. The Crystal Diode

When an N-type and a P-type semiconductors (for example, germanium) are joined together, a semiconductor (or crystal) diode (rectifier) is made (Fig. 16-3).

Due to the high concentration of electrons in the N-type germanium as compared with the P-type, free electrons diffuse from the former into the latter. In a like manner, free holes diffuse from the P-type into the N-type germanium. The thin boundary region soon becomes devoid of holes on the *P* side and of electrons on the *N* side. The result is an accumulation of positive charges at the boundary on the *N* side (a positive space charge) and of negative charges on the *P* side (a negative space charge). Now no more free electrons on the *N* side can go over to the *P* side because of the opposing forces of its negative ions, and no more free holes on the *P* side can go over to the *N* side because of the

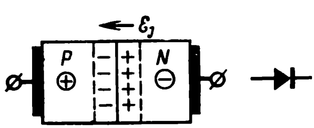


Fig. 16-3. Semiconductor diode (rectifier) and its diagram symbol

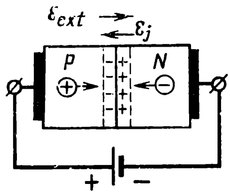


Fig. 16-4. Forward-biased rectifier

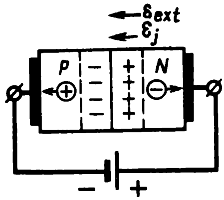


Fig. 16-5. Reverse-biased rectifier

opposing forces of its positive ions. Thus, a potential barrier is formed, and the potential difference across this barrier gives rise to an electric field, $\mathscr{E}_j$, which stops further diffusion. The thin boundary layer devoid of majority carriers and having a high resistance is called *depletion* (*or barrier*) *layer*, or the *P-N transition region.*

The diffusion of majority carriers across the boundary layer is accompanied by the reverse motion of minority carriers (holes from the *N* region into the *P* region and of electrons from the *P* region into the *N* region), caused by the electric field established by the junction potential difference. This reverse flow of minority carriers directed against the diffusion current is called *minority-carrier*, or *field-conduction*, *current.* This current depends on the temperature of semiconductors and is often called *thermal current.* The currents due to diffusion and field conduction are equal in magnitude and opposite in polarity, so the total charge crossing the transition region per unit time is zero, as it should be, when no external field is applied.

Let us connect the "+" terminal of a source to the metal electrode on the *P* side and the "—" terminal to that on the *N* side of a *P-N* transition region, as shown in Fig. 16-4. This will establish an external electric field, $\mathscr{E}_{ext}$, opposing the field of the *P-N* transition region. This field will drive electrons and holes toward the boundary layer. As this takes place, the number of majority carriers in the barrier region rises, the space charge decreases, and so do the magnitude of the potential barrier and the resistance of the depletion layer. This is called *forward biasing*, and the respective current is called the *forward current*, I_f; it is considerable even at a low voltage applied.

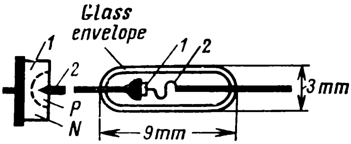


Fig. 16-6. Germanium point-contact rectifier

Let us reverse the polarity of the external source (Fig. 16-5). Now the external field acts in the same direction with the field of the *P-N* transition region, and the resultant field thus grows in intensity, so it will prevent majority carriers from passing through the depletion layer still more. This is *reverse biasing*, and associated with it is the *reverse current*, I_r, which is caused by minority carriers. Since their concentration strongly depends on temperature it is also called the thermal current. This current is rather low and may be neglected in some cases.

Thus, a *P-N* transition region possesses pronounced unidirectional conduction, and it acts as a rectifier. The diagram symbol of a crystal (or semiconductor) rectifier is given in Fig. 16-3.

The ratio of the forward current to the reverse current at the same voltage is called the *rectification factor*:

$$k_r = I_f/I_r \tag{16-1}$$

16-4. Germanium and Silicon Diodes

As already noted, a device utilizing unidirectional conduction across a *P-N* transition region is called a crystal diode or crystal rectifier. Such a device has one *P-N* transition region and two terminals or leads.

According to their design, crystal diodes can be classed into the *point-contact type* and the *junction type*.

A point-contact germanium diode (Fig. 16-6) consists of a glass (or metal-glass) envelope or case 3 mm in diameter and 9 mm in length into which two leads are sealed. One of the leads gives support to an N-type germanium crystal, *1*, and the other carries a pointed wire (called a catwhisker), *2*, made of tungsten or gold. The tip of the wire is welded to the crystal surface by "shots" of current. The heavy current melts the semiconductor material around the point

of contact, so that after forming (as it is called), a small-diameter hemispherical *P-N* transition region is produced (Fig. 16-6). The maximum forward current of this rectifier is 16 mA and the peak-inverse voltage is 50 V. Due to the small surface of the *P-N* transition region, the diode has a capacitance of about 1 pF and low power dissipation.

A germanium junction diode (Fig. 16-7*a*) consists of a wafer, *1*, made of arsenic- or antimony-doped N-type germanium, and a dot of indium, *2*, which is alloyed on a side of the wafer at about 500°C. In the molten state, the atoms of indium diffuse into the germanium, thereby forming a layer of P-type germanium, *2a* (Fig. 16-7*a*). Thus, a *P-N* junction is formed at the boundary of these two regions.

Figure 16-7*b* shows a sketch of a junction germanium diode. The germanium crystal, *1*, is carried by a crystal-holder, *3*, to which the bottom lead, *4*, is welded. The top lead, *4*, is connected by an internal lead, *5*, to the indium electrode, *2*. The metal envelope, *6*, is welded to the crystal-holder, *3*, and a glass insulator, *7*. Figure 16-8 is the volt-ampere characteristic of a germanium rectifier (curve *1*).

Germanium diodes can handle current densities up to 100 A/cm^2 at a forward voltage drop of up to 0.8 V. The maximum peak-inverse voltage is $V_{p\text{-}i} = 400$ V. They can operate in the temperature range from -60°C to $+75$°C.

Silicon rectifier diodes are made by alloying a dot of aluminium into a wafer of N-type silicon. These diodes can handle current densities up to 200 A/cm^2 at a forward voltage drop of up to 1-1.2 V. The peak-forward current $I_{p\text{-}f}$ is over 1000 A and the peak-inverse voltage $V_{p\text{-}i}$ exceeds 800 V. The temperature range extends from -60°C to $+150$°C. The volt-ampere characteristic of a silicon diode is given in Fig. 16-8 (curve *2*). Figure 16-9 shows a Soviet-made type BK-100 air-cooled silicon diode intended for a current of 100 A.

An increase in temperature boosts the generation of electron-hole pairs in semiconductors, and their intrinsic conductivity increases, as do the forward and, still more, the reverse current of a crystal diode. In germanium and silicon diodes, the reverse current rises about 2 or 2.5 times per

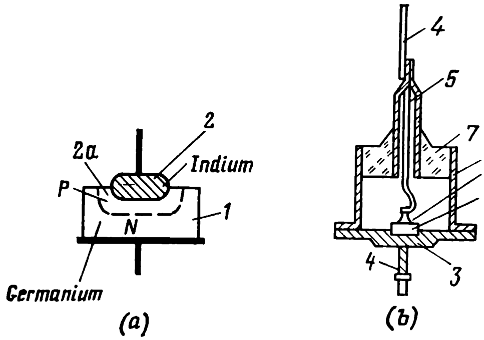


Fig. 16-7. Soviet-made type Д-7 germanium **junction** rectifier

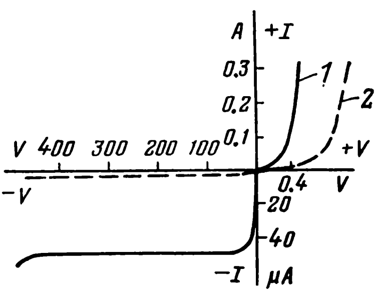


Fig. 16-8. Volt-ampere characteristics of (*1*) a germanium diode and (*2*) a silicon **diode**

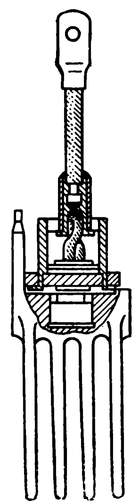


Fig. 16-9. Soviet-made type BK-100 silicon diode

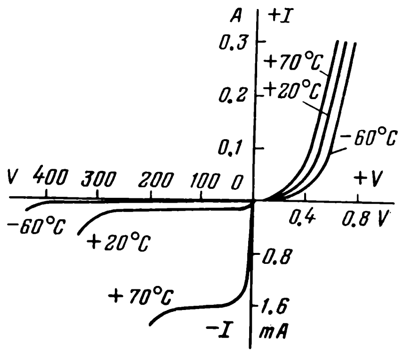


Fig. 16-10. Effect of temperature variations on the volt-ampere characteristic of a germanium diode

ten degrees of temperature increase. The effect of temperature variations on the volt-ampere characteristic is shown in Fig. 16-10.

The main ratings and parameters of crystal diodes are: peak-forward current $I_{p\text{-}f}$, forward voltage drop V_f at $I_{p\text{-}f}$, peak-inverse voltage $V_{p\text{-}i}$, peak-reverse current $I_{p\text{-}r}$ at $V_{p\text{-}i}$, maximum power dissipation $P_{d\text{-max}}$, junction capacitance C; maximum allowable frequency f_{max}, and operating temperature range.

16-5. Copper-Oxide and Selenium Diodes

A copper-oxide rectifier (Fig. 16-11) consists of a copper disc, *1*, with a layer of cuprous oxide, *2*, formed on it. For better contact, it is covered by a snugly fitting lead disc, *3*, followed by a large-diameter brass disc, *4*, which serves as a heat-sink. The layer of cuprous oxide is obtained by a thermal treatment of copper in an oxygen atmosphere. The outer layer of cuprous oxide, *2'*, obtained in the atmosphere with an excess of oxygen, has hole conduction; the layer of cuprous oxide, *2"*, next to the copper baseplate, is produced in the atmosphere with a deficit of oxygen and has electron conduction. The two layers of cuprous oxide form a *P-N* junction.

The peak-inverse voltage across this rectifier is not more than 10 V because it can break down at a peak-inverse voltage of 20 or 30 V. For higher voltages, they are put

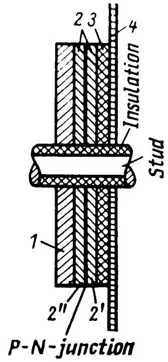


Fig. 16-11. Copper-oxide rectifier

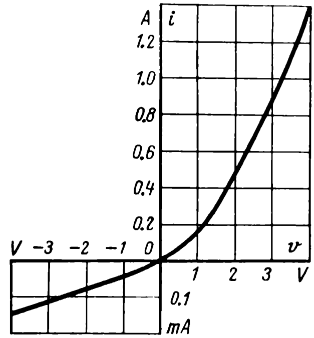


Fig. 16-12. Volt-ampere characteristic of a copper-oxide rectifier

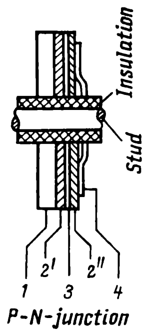


Fig. 16-13. Selenium rectifier

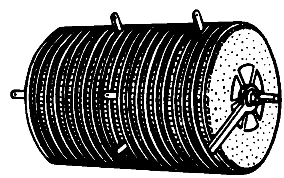


Fig. 16-14. Stack of selenium rectifiers

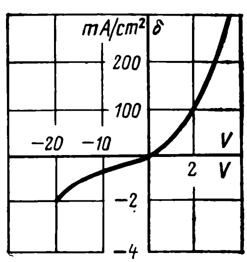


Fig. 16-15. Volt-ampere characteristic of a selenium rectifier

together on a stud to form a rectifying stack or pile. For better cooling, they are equipped with radiator washers to keep the temperature under +60°C, because otherwise the rectifier may lose its rectifying properties. Copper-oxide rectifiers can handle current densities up to 0.1 A/cm^2. The volt-ampere characteristic of a copper-oxide rectifier is given in Fig. 16-12.

A selenium rectifier (Fig. 16-13) consists of an aluminium or steel base plate, *1*, which gives support to a coating of crystalline selenium, *2″*, having hole conduction and serving as an electrode. The other electrode (known as the counter electrode), *2′*, is a metallic film of a cadmium and tin alloy. The atoms of cadmium diffuse into the selenium and form an N-type layer. In this way, the crystalline and cadmium-doped selenium compose a barrier layer, *3*. The counter electrode, *2′*, is held down by a spring washer, *4*.

The peak inverse voltage across a selenium rectifier is from 20 to 40 V; should the voltage rise up to 60 or 80 V, the rectifier might break down. The operating temperature should not exceed +75°C, and the current density should not be over 0.1-0.2 A/cm^2. Figure 16-14 shows a stack of selenium rectifiers. The volt-ampere characteristic of a selenium rectifier is given in Fig. 16-15.

16-6. Application of Crystal Rectifiers

According to their purpose, diodes can be divided into two groups: diodes intended to rectify power- and high-frequency current, and those for conversion of r.f. signals into audio frequency signals, that is, for detection.

In addition to germanium and silicon rectifiers of which the latter prevail in heavy-current installations, the earlier types of selenium and copper-oxide rectifiers are still manufactured.

Although they can handle lower current densities and will stand up to lower inverse voltages, these rectifiers are easy and inexpensive to make, so they find application in a number of fields. Selenium rectifiers are used in battery chargers, electroplating and electrolysis equipment operating at low voltages and high currents, and also at

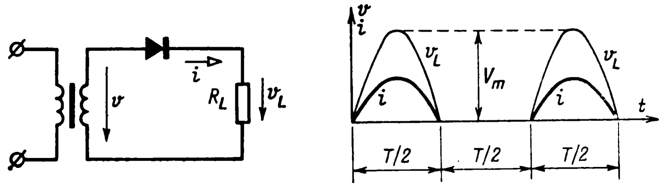


Fig. 16-16. Half-wave rectifier circuit and its load current and voltage waveforms

low currents and high voltages where rectifiers can be connected in series.

Copper-oxide rectifiers are used in the instrumentation field owing to the stability of their parameters, and also in electrolyzers operating at 4-6 V.

One of the main applications for crystal rectifiers is the rectification of alternating current, that is, its conversion into direct (pulsating) current.

Figure 16-16 shows an elementary rectifier circuit. In this circuit, the rectified current i flows through a crystal diode and load R_L only during the positive half-cycles of the alternating voltage v supplied by the secondary winding of a transformer. This current produces a rectified voltage v_L across the load. During the negative half-cycles, the current does not flow through the load because the diode is turned off (nonconducting), and the peak inverse voltage across the diode is equal to the voltage amplitude, V_m.

A major drawback of this circuit is the ripple in the load current and voltage. The ripple is reduced by smoothing (ripple) filters. Rectifiers and filters will be discussed in greater detail in Chapter 18.

If it is necessary to obtain a forward current which exceeds the safe value for one diode, several diodes of the same type are connected in parallel (Fig. 16-17). To eliminate the difference in the currents flowing through the diodes, they are connected in series with low-value equalizing resistors. Owing to this arrangement, the parallel branches have almost the same resistance, although the diodes may differ widely in resistance.

If it is desired to obtain an inverse voltage which exceeds the peak inverse voltage for one rectifier, diodes should be

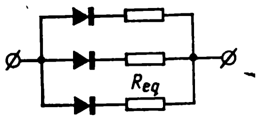


Fig. 16-17. Diodes in parallel

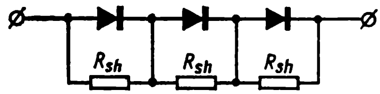


Fig. 16-18. Diodes in series

connected in series (Fig. 16-18). To secure equal division of the inverse voltage among the series-connected rectifiers with different reverse resistances, they are shunted with resistors R_{sh}, which are by about an order of magnitude lower than the reverse resistance of the rectifiers.

Owing to the stability of their parameters and considerable slope of their volt-ampere characteristics, copper-oxide and selenium rectifiers can be series-connected without equalizing resistors.

16-7. Marking of Crystal Diodes

In the Soviet Union, the designations of low-power crystal diodes (to GOST 10862-64) consist of four elements (Russian letters throughout).

The first element is a letter or numeral which indicates the material: Г or 1 stand for germanium, К or 2 for silicon, and А or 3 for gallium arsenide.

The second element is a letter which characterizes the type of the device: Д stands for diodes, Ц for rectifying stacks or piles, and С for Zener diodes.

The third element is the number which indicates the purpose or performance of the device. The numbers from 101 to 199, from 201 to 299 and from 301 to 399 stand for rectifier diodes with average forward currents of up to 0.3 A, 0.3 to 10 A, and over 10 A, respectively. The numbers from 401 to 499 stand for universal diodes, and from 501 to 599 for pulse diodes.

The fourth element is a letter which characterizes the modification of the given group of devices.

Power diodes are designated in a different way.

16-8. The Silicon Zener Diode

A silicon Zener diode is a good direct voltage regulator and a reliable voltage reference (this is the reason why it is also called a silicon VR diode). From conventional silicon diodes, the Zener diode differs by an increased concentration of charge carriers.

For its operation the silicon Zener diode utilizes the knee, or sharp break, in its volt-ampere characteristic, which corresponds to reverse current and reverse voltage and runs parallel to the current axis (the solid line in Fig. 16-19).

The maximum current of the Zener diode, $I_{st,\,max}$, is determined by the maximum power dissipation

$$I_{st,\,max} = P_{max}/V_r \tag{16-2}$$

and is limited by ballast resistor R_b.

The maximum current of various types of the Zener diodes ranges from 20 mA to 2 A. The rated voltage runs from 6 to 400 V. The dynamic resistance within the effective range is from 1 ohm to 70 ohms, depending on the type.

As is seen from the circuit diagram of a voltage stabilizer using a silicon Zener diode (Fig. 16-20), the load is connected

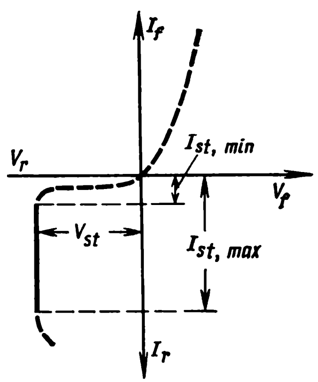


Fig. 16-19. Volt-ampere characteristic of a silicon Zener diode

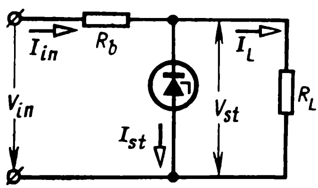


Fig. 16-20. Voltage stabilizer using a silicon Zener diode

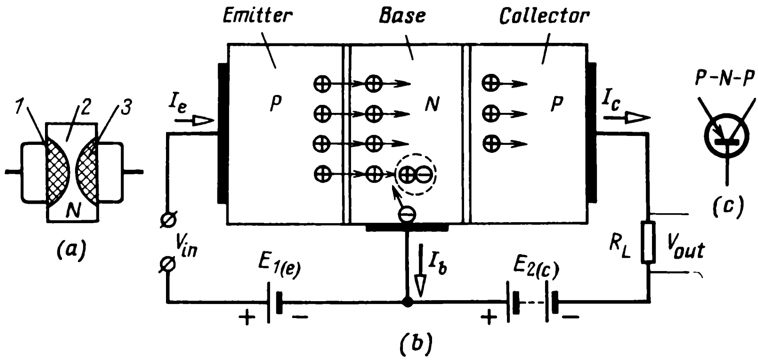


Fig. 16-21. Connection of a P-N-P transistor in a circuit

in parallel with the Zener diode and the ballast resistor R_b is connected in the common part of the circuit. In this circuit, the Zener diode is reverse-biased.

16-9. Transistors

A semiconductor device with two P-N junctions and three electrodes is called a transistor*. It can be used for power amplification and generation of electric oscillations.

A transistor (Fig. 16-21*a*) consists of a thin N-type germanium wafer sandwiched between two pellets of indium. Indium atoms diffuse into the germanium and form two regions (*1* and *3*) with hole (P-type) conduction. The N region is made very thin—from several to a few tens of micrometres. The regions separated by the P-N junctions (Fig. 16-21*b* and *c*) are called the *emitter region* (E), the *base region* (B) and the *collector region* (C), to each of which is made the respectively termed electrode and lead (or terminal).

To begin with, assume that the emitter-base circuit is open, its current is zero, $I_e = 0$, and a reverse voltage, E_2, of about ten volts is applied between the collector and base. In the circumstances, a small reverse (thermal) collec-

* This is the most commonly used type of transistor. There also exist transistors with four, five and more P-N junctions.— *Translator's note.*

tor current, I_{c0}, due to minority charges flows in the collector circuit. This current is one of the transistor parameters. The lower this current, the better the semiconductor material used.

Now let a direct voltage E_1 (of about several volts) be applied between the emitter and base.

The concentration of impurity atoms and holes in the emitter region is much greater than the concentration of impurity atoms and free electrons in the base region. The emitter-to-base voltage, E_1, biases the junction in the forward direction. As the forward resistance of the *P-N* junction is low, the emitter current, which is mainly due to holes, is relatively high even at a low value of E_1. The minority carriers are said to be injected from the emitter into the base. In the base, an insignificant fraction of the holes recombines with free electrons, and their disappearance is made up for by new electrons coming from the external circuit, thereby giving rise to the base current I_b. Diffusion causes a greater proportion of holes to keep moving in the base, reach the base-collector junction, and pass through the *P-N* junction to the collector electrode under the action of the electric field established by the source E_2. This shows that the collector is intended to extract minority carriers coming from the base region. Thus, the current $I_c = I_e - I_b$ arising in the base-collector circuit is of the same order of magnitude as the current in the emitter-base circuit. The ratio of an incremental change ΔI_c in the collector current to an incremental change ΔI_e in the emitter current with the collector voltage held constant is called the *alpha current gain*:

$$\alpha = k_i = \Delta I_c/\Delta I_e \quad \text{at} \quad V_c = \text{constant} \qquad (16\text{-}3)$$

The alpha current gain is always less than unity and ranges from 0.9 to 0.995.

Figure 16-22 shows a sketch form of a germanium junction transistor. The base region is a wafer, *10*, of N-type germanium. It is carried by a support, *9*, connected to a lead, *2*. On either side of the base is an indium electrode, *8* and *11*. When indium is fused into the germanium, P-type regions are formed between each of these electrodes and the germanium base. The triode is enclosed in a metal case, *5* and *6*.

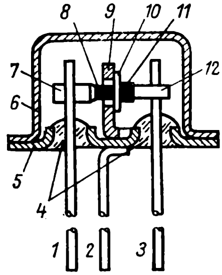


Fig. 16-22. Soviet-made type П-13 germanium junction transistor

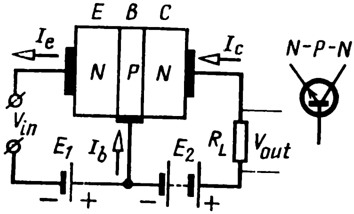


Fig. 16-23. Connection of an *N-P-N* transistor in a circuit

The leads from the emitter (*7* and *1*) and the collector (*12* and *3*) are insulated from the case by glass bushings, *4*.

In addition to *P-N-P* transistors, use is also made of *N-P-N* transistors (Fig. 16-23) which operate in the same way. In a *N-P-N* transistor, the voltage across the emitter and base forces the *N* region to emit electrons to the *P* region. The polarity of the sources E_1 and E_2 should be reversed as compared with that of the respective sources for a *P-N-P* transistor.

In the circuits considered above (Fig. 16-21 and Fig. 16-23), the base is common to the emitter and collector circuits, so these are common-base circuits.

16-10. Application of Transistors

(a) Signal Amplification

When a transistor operates as a signal amplifier, an input alternating voltage v_{in} (the signal to be amplified) is applied between the emitter and base in series with the bias voltage source E_1 (Fig. 16-24*a*) or E_b (Fig. 16-25*b*), and the load resistor R_L the voltage across which is the amplified signal is connected in series with a source E_2, whose positive terminal is connected to the emitter. This is a common-emitter circuit, because the emitter is common to the base and collector circuits.

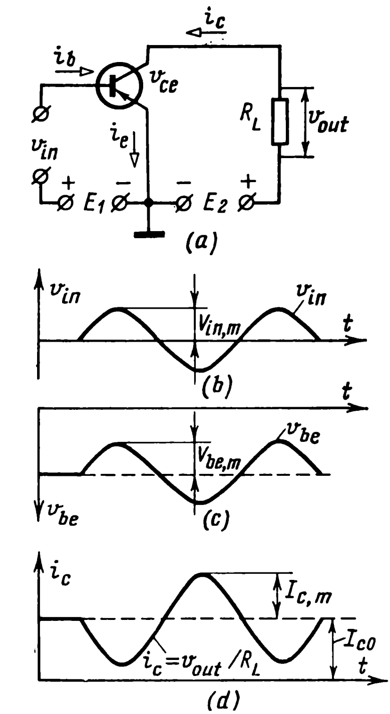


Fig. 16-24. (*a*) Connection of a transistor and (*b*, *c*, *d*) its voltage and current waveforms

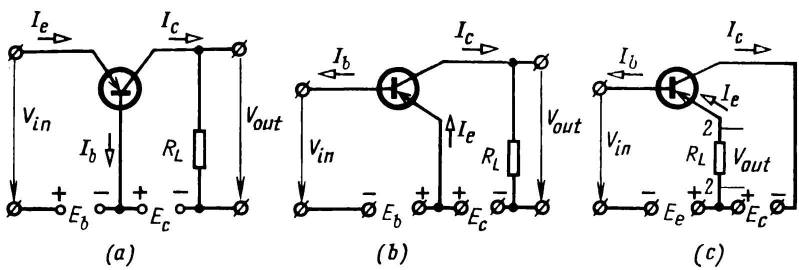


Fig. 16-25. Transistor circuit configurations
(*a*) common-base; (*b*) common-emitter; (c) common-collector

If a negative bias voltage, that is, the forward voltage for the emitter junction, is applied to the base, a current will flow in the base circuit, and also in the collector circuit. This will give rise to voltages across the load resistor R_L and the transistor resistance $r_{c0} + r_e =$ (approx.) r_{c0}. In the presence of sinusoidal input signals (Fig. 16-24*b*), the currents in the base and collector circuits will change and so will the resistance of the collector junction; the result will be a redistribution of voltage between r_{c0} and R_L. A rise in the collector current will cause the voltage between the collector and emitter to go down, and that across the R_L to go up. In the circumstances, the a.c. component of the load voltage may be tens of times the input voltage because $I_c \gg I_b$ and $E_2 > E_1$.

Figure 16-24*c* shows the direct and alternating components of the voltage between the base and emitter, V_{be}. Figure 16-24*d* gives the direct and alternating components of the collector current. The alternating component of the output voltage $v_{out} = i_c R_L$ is proportional to the current, so, if plotted to another scale, the current curve is the curve of the output voltage v_{out}.

(b) Configurations of Transistor Circuits

The three common transistor configurations are known as (1) common- or grounded-base (Fig. 16-25*a*); (2) common- or grounded-emitter (Fig. 16-25*b*); and (3) common- or grounded-collector (Fig. 16-25*c*). The name of each circuit shows which electrode of a transistor is common to the input and output circuits.

Although these configurations differ in properties, the principle of amplification is the same.

The common-base configuration has already been considered. In this configuration, the input voltage source is connected in the emitter-base circuit, and the load and power source are connected in the collector-base circuit. The input resistance of the common-base circuit is low, ranging from several to a few tens of ohms. This is because the emitter junction is biased in the forward direction. In contrast, the output resistance of the circuit is high,

being from hundreds of kilohms to several megohms, because the collector is biased in the reverse direction.

The low resistance of the common-base circuit is a major disadvantage which limits its application in amplifiers. It is mainly used in transformer coupled amplifiers.

In such a circuit, all of the emitter current flows through the input voltage source, so there is no current amplification. As already noted, the alpha current gain is at best 0.9 to 0.995. The voltage and power gain in this circuit may be as high as several hundred.

Figure 16-25*b* shows the common-emitter configuration. In this case, the input voltage source is connected in the base-emitter circuit, and the load resistor R_L and power source are connected in the emitter-collector circuit, so that the emitter is common to the input and output circuits. The input resistance of the common-emitter circuit exceeds that of the common-base circuit and is equal to several hundred ohms. This is because the input current is the base current which is considerably lower than the emitter and collector currents. The common-emitter circuit has a high output resistance which runs into handreds of kilohms.

As is seen from Fig. 16-24, voltage amplification in the common-emitter circuit, is accompanied by a phase reversal of the output voltage with respect to the input voltage.

The current gain of a common-emitter circuit is designated by the letter β and is defined as the ratio of an incremental change in the collector current ΔI_c to an incremental change in the base current ΔI_b with the collector-to-base voltage held constant:

$$\beta = \Delta I_c/\Delta I_b, \quad V_c \text{ being constant} \tag{16-4}$$

Noting that $I_e = I_c + I_b$, we may write:

$$\beta = \frac{\Delta I_c}{\Delta I_e - \Delta I_c} = \frac{\Delta I_c/\Delta I_e}{(\Delta I_e/\Delta I_e) - (\Delta I_c/\Delta I_e)} = \frac{\alpha}{1-\alpha} \tag{16-5}$$

For this circuit, the value of β ranges from 10 to 200, and the voltage gain is of the same order as that of the common-base circuit, that is, several hundred. The power gain is

$$k_p = k_i k_v = \beta k_v$$

and runs into several thousands which is many times the power gain of the common-base circuit.

Owing to these advantages, this configuration is most commonly used in amplifiers.

In the common-collector configuration (Fig. 16-25*c*), the input voltage source is connected in the base circuit, and the load resistor R_L in the emitter circuit. The input current is the base current, and the output current is the emitter current. For this configuration, the current gain is

$$k_i = \Delta I_e/\Delta I_b = \Delta I_e/(\Delta I_e - \Delta I_c) = \frac{1}{1-\Delta I_c/\Delta I_e}$$

$$= \frac{1}{1-\alpha} \qquad (16\text{-}6)$$

The input resistance of the common-collector circuit is high (up to tens of kilohms) and the output resistance is low (a few kilohms). The voltage gain of the common-collector stage is from 0.9 to 0.95, and the power gain is a few tens.

The common-collector configuration is used rather seldom. It mainly serves to match amplifier stages of power sources and loads to amplifiers. As its voltage gain k_v is close to unity, this circuit is often called an emitter-follower.

(c) Transistor Characteristics

The transistor characteristics represent the relationship between currents in and voltages across the input and output circuits.

The input characteristics of a *P-N-P* transistor connected in a common-base circuit are given in Fig. 16-26:

$$I_e = f(V_{eb}), \quad V_{cb} \text{ being held constant}$$

When the voltage between the emitter and base V_{eb} is low, the emitter current, I_e, first rises slowly due to the high resistance of the *P-N* junction, then it increases more rapidly.

In Fig. 16-26*a*, one characteristic is plotted at the zero collector-to-base voltage ($V_{cb} = 0$), and the other at $V_{cb} = -15$ V. As V_{cb} becomes more negative, the input characteristic shifts to the left, because the emitter current slightly increases under the action of the field that V_{cb} establishes at the emitter junction.

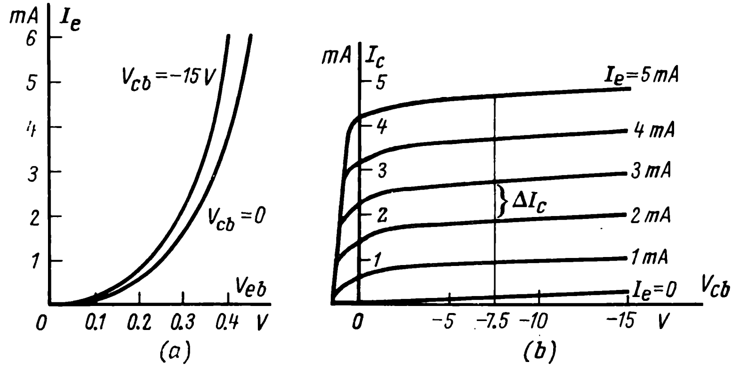


Fig. 16-26. Volt-ampere characteristics of a *P-N-P* transistor connected in a common-base circuit

Figure 16-26*b*, shows the output characteristics:

$$I_c = f(V_{cb}), \quad V_e \text{ being held constant}$$

These characteristics show that the collector-base voltage V_{cb} has an insignificant effect on the collector current, I_c, which mainly depends on the number of holes injected into the base, that is, on the emitter current, I_e.

The output characteristics make it possible to determine the alpha current gain (see Fig. 16-3): α=(approx.) $\Delta I_c/\Delta I_e$, where ΔI_c is the difference in ordinate between two characteristics (for example, 2 and 3 mA in Fig. 16-26*b*) at the same abscissa (for example, —7.5 V), and ΔI_e is the difference between the emitter currents at which the characteristics were plotted (for example, $\Delta I_e = 3 - 2 = 1$ mA).

The input and output characteristics of a *P-N-P* transistor connected in a common-emitter circuit are given in Fig. 16-27.

The input characteristics (Fig. 16-27*a*) represent the relationship between the base current I_b and the base-to-emitter voltage V_{be}, with V_{ce} held constant:

$$I_b = f(V_{be}) \text{ at } V_{ce} \text{ held constant}$$

At low values of the V_{be}, the base current I_b increases first slowly, then more rapidly with rising V_{be}, until it reaches a certain value where it no longer increases. The

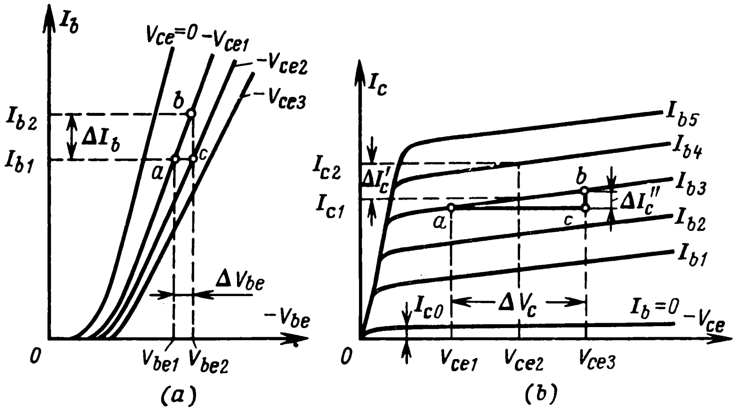


Fig. 16-27. Static characteristics of a *P-N-P* transistor connected in a common-emitter circuit

slope of the linear portions of the characteristics is not the same for different output voltages, V_{ce}.

The input characteristics shown in Fig. 16-27*b* are

$$I_c = f(V_{ce}), \quad I_b \text{ being constant}$$

Within the effective range, the output characteristics of a *P-N-P* transistor connected in a common-emitter circuit rise more steeply than those of a transistor connected in a common-base circuit. This is the result of the effect that the collector voltage has on the injection of holes into the base.

For a *P-N-P* transistor connected in a common-collector circuit, one usually uses the same characteristics as for the common-emitter circuit.

The parameters used to characterize the properties of transistors can be divided into primary and secondary.

The primary parameters include:

(1) emitter junction resistance, r_e (tens of ohms);

(2) base resistance, r_b (hundreds of ohms);

(3) collector junction resistance, r_c (tens of kilohms);

(4) current gain (alpha or beta according to the circuit configuration).

The secondary parameters determine the relationship between incremental changes in currents in and voltages across a transistor. They vary according to the circuit configuration. The most widely used are the *h*- (or hybrid) parameters. They can be determined from the static characteristics. For the common-emitter circuit (Fig. 16-25*b*), the *h*-parameters can be found from the characteristics shown in Fig. 16-27.

The *h*-parameters include the following quantities.

1. Input impedance:

$$h_{11} = \Delta V_{be}/\Delta I_b = (V_{be2} - V_{be1})/(I_{b2} - I_{b1}), \quad V_{ce1} = \text{constant} \tag{16-7}$$

2. Voltage feedback ratio:

$$h_{12} = \Delta V_{be}/\Delta V_c = (V_{be2} - V_{be1})/(V_{ce2} - V_{ce1}), \quad I_{b1} = \text{constant} \tag{16-8}$$

3. Forward current transfer ratio:

$$h_{21} = \Delta I'_c/\Delta I_b = (I_{c2} - I_{c1})/(I_{b4} - I_{b3}), \quad V_{ce2} = \text{constant} \tag{16-9}$$

4. Output admittance:

$$h_{22} = \Delta I''_c/\Delta V_c = \Delta I''_c/(V_{ce3} - V_{ce1}), \quad I_{b3} = \text{constant} \tag{16-10}$$

Transistors offer a number of advantages over vacuum-tube devices. As they have no filament, their circuit is simpler, they have greater mechanical strength and longer service life, instantaneous readiness for operation, smaller size and mass, low supply voltage and a higher efficiency.

Among the disadvantages of transistors are the dependence on ambient temperature, considerable difference between input and output impedances, low power output, overload sensitivity, and spread in parameters, that is, a marked difference in parameters between individual transistors of the same type.

16-11. Type Designations of Transistors

Transistors can be classified according to their electric parameters and performance. For example, in accordance with power dissipation, they can be divided into low-power

(usually up to 0.3 W), medium-power (from 0.3 to 3 W) and high-power (over 3 W) transistors. By signal frequency, they can be classed into low-frequency (up to 10 MHz), medium-frequency (from 10 to 100 MHz) and high-frequency (over 100 MHz) transistors.

In the Soviet Union, transistors are designated as follows (to GOST 10862-64). The first element is a letter or a numeral which indicates the material (Russian letters throughout): 1 or Г stands for germanium, 2 or K for silicon, and 3 or A for gallium arsenide. The second element is the letter T which stands for transistor. The third element is a three-digit number in which the first digit indicates the frequency and power dissipation group and the remaining two the ordinal type number, namely:

101 to 199: low-power, low-frequency transistors;
201 to 299: low-power, medium-frequency transistors;
301 to 399: low-power, high-frequency transistors;
401 to 499: medium-power, low-frequency transistors;
501 to 599: medium-power, medium-frequency transistors;
601 to 699: medium-power, high-frequency transistors;
701 to 799: high-power, low-frequency transistors;
801 to 899: high-power, medium-frequency transistors;
901 to 999: high-power, high-frequency transistors.

Sometimes there may be a range of transistor modifications of a basic type, differing only in some parameters. In such a case, they have a letter at the end of the designation to identify the modification. For example, ГТ308А stands for a low-power, high-frequency germanium transistors, modification A; КТ312В for a low-power, high-frequency silicon transistor, modification B.

16-12. Thyristors

A thyristor is a four-layer (PNPN), three-junction, three-terminal semiconductor device which exhibits switching action. Referring to Fig. 16-28, its four layers, P_1, N_1, P_2 and N_2, form three junctions, J_1, J_2 and J_3 to which are made three terminals: anode (A), cathode (K), and gate (G).

One of the oldest and most commonly used members in the thyristor family is the silicon-controlled rectifier (SCR),

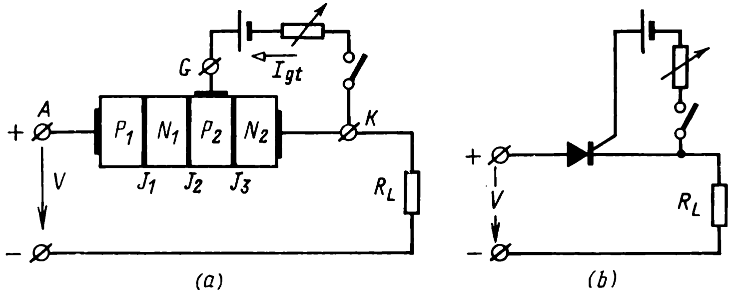


Fig. 16-28. (*a*) Construction and (*b*) circuit diagram of a silicon controlled rectifier

so the discussion that follows will mainly be concerned with the SCR and its operation.

As in a transistor, the outer layers, P_1 and N_1, are called the emitter and collector regions, and the layer connected to the gate is called the base region.

When the supply voltage V is low and the gate circuit is open, the two outer junctions, J_1 and J_3, are biased in the forward direction, and the middle junction J_2, is biased in the reverse direction. Since J_2 offers a high resistance as compared with J_1 and J_3, it drops a considerable part of the supply voltage, and the current in the circuit is low.

As the voltage across the SCR is raised, the current in it rises insignificantly, because it is limited by the high resistance of J_2 (Fig. 16-29, curve *Oa*). At what is known as the *forward breakover voltage*, V_{bo}, the field at J_2 grows strong enough for ionization to take place, that is, for free electrons and holes to be produced. This goes on in a cumulative manner until an *avalanche breakdown* takes place at J_2. The current in the thyristor rises to a value known as the *breakover current*, I_{bo}, the voltage across the junction J_2 rapidly falls (to about 1V) because its resistance becomes negligible, and the SCR turns on. The power dissipated at J_2 is low, and the avalanche breakdown taking place at normal voltage will not cause irreversible changes in the SCR structure.

The "ON" state of the device is represented by region *ab* (Fig. 16-29). The adjacent region *bc* is similar to the normal volt-ampere characteristic of a silicon rectifier. The SCR

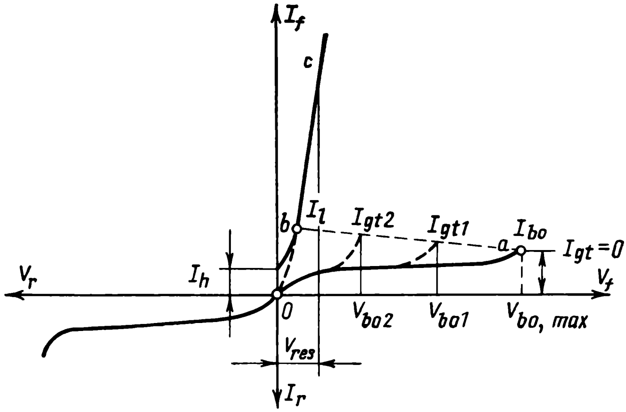


Fig. 16-29. Volt-ampere characteristics of a silicon-controlled rectifier

remains in the ON state as long as the current in the thyristor is high enough to produce charge carriers in J_2. When the current decreases to lower than what is known as the *holding current*, I_h, the SCR will turn off.

If we connect the "+" terminal of an auxiliary battery to the gate, the *gate trigger current* I_{gt} will inject an additional number of charge carriers, that is, electrons, into the base region. This brings down the breakover voltage of J_2. Raising the gate trigger current lowers the voltage at which the thyristor is rendered conducting. As the current applied to the gate is raised to what is known as the *latching current* I_l, the SRC will remain on and operate as an uncontrolled rectifier (Fig. 16-29, region *Obc* of the volt-ampere characteristic).

To sum up, a thyristor can be turned on (rendered conducting) either by applying the breakover voltage V_{bo} at which the avalanche breakdown occurs, or by applying the gate trigger current I_{gt} so that additional charge carriers are injected into the base region. As it takes about 10 μs to turn on a thyristor and the gate has no control over its operation after it has been turned on, a short gate trigger pulse will be enough to cause the SCR to turn on.

When a reverse voltage, V_r, is applied to its terminals, its junctions J_1 and J_3 are biased in the reverse direction,

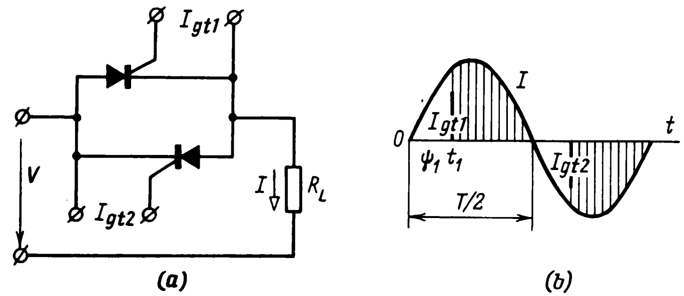


Fig. 16-30. (*a*) Two SCRs connected in parallel front-to-back and (*b*) load current waveform

and block the SCR irrespective of the voltage at the gate. Now its volt-ampere characteristic does not differ from that of a conventional rectifier (Fig. 16-29).

In addition to triode thyristors, there are also diode thyristors. Both types are often used as static switches. They have two stable states; in one state, the resistance of the device is high (r = approx. ∞), so that the circuit is open, and, in the other, the resistance is low (r = approx. 0), so that the circuit is closed.

Figure 16-30 shows two SCRs connected in parallel front-to-back and placed in a circuit driven by a sinewave voltage and containing a resistive load R_L. If we apply gate trigger pulses, I_{gt1}, to the first SCR at the beginning of each cycle ($\alpha = 0$), the current in this branch will flow during the whole half-cycle, $T/2$. If we apply gate trigger pulses with a time shift, $t_1 = \alpha_1/\omega$, the current will flow in this branch only during a part of the half-cycle, between t_1 and $T/2$ (Fig. 16-30*b*). The same will take place in the second branch, if the time shift between gate trigger pulses I_{gt1} and I_{gt2} is a half-cycle (Fig. 16-30*b*). Thus, when the gate trigger pulses I_{gt1} and I_{gt2} are applied at different instants in a cycle (at different firing angles α), the duration of current flow can be varied during each half-cycle and, therefore the rms load current defined as the half-cycle-averaged ordinate of the shaded area (Fig. 16-30), can be controlled without resort to any devices with movable controls.

The main ratings of SCRs are:

(1) forward breakover voltage, V_{bo}, defined as the anode voltage at which the SCR turns on, with the gate circuit open;

(2) residual voltage, V_{res}, defined as the forward voltage across the device in the "ON" state at the nominal current;

(3) reverse current, I_r, defined as that existing at a specified voltage;

(4) gate trigger current, I_{gt}, defined as the minimum d.c. gate current required to cause the SCR to switch from the nonconducting to the conducting state;

(5) maximum current, $I_{f\text{-max}}$, defined, as the maximum safe current in the "ON" state;

(6) holding current, I_h, defined as the minimum current below which the SCR will not stay on;

(7) breakover current, I_{bo}, defined as the current corresponding to the breakover voltage;

(8) firing time, τ_f, defined as the time interval between the instant of a gate trigger pulse and the instant when the voltage across the SCR reaches 10% of the initial value;

(9) turn-off time, $\tau_{t\text{-}o}$, defined as the minimum time interval during which the reverse voltage should be applied to the SCR to turn it off.

Chapter Seventeen Photoelectric Devices

17-1. Photocells

A *photocell* (or *phototube*) is a vacuum, gas-filled, semiconductor or any other device whose electrical properties (current, resistance or emf) vary with the intensity of light that strikes its active material

According to the medium through which electrons move, photocells can be divided into three classes:

— *vacuum-tube photocells* in which electrons move in a vacuum;

— *gas-filled photocells* in which electrons move in and ionize a rarefied gas;

— *semiconductor photocells* in which electrons liberated by light either increase their conductivity or produce an emf.

Vacuum and gas-filled phototubes utilize *photoelectric emission*, also known as the *outer* or *external photoeffect*.

As explained in Sec. 13-4, photoelectric emission consists in that incident light imparts additional energy to the electrons of the material so that they escape from the surface into a vacuum or rarefied gas.

Photo-conductive cells utilize the *photoconductive effect*, also known as the *internal* or *inner photoelectric effect*.

The photoconductive effect consists in that incident light imparts additional energy to some electrons, ionizes some atoms, and causes the production of new charge carriers — electrons and holes, so that the resistance of the material goes down.

Semiconductor photocells, such as photodiodes and phototransistors, utilize the *photovoltaic effect*. In this case,

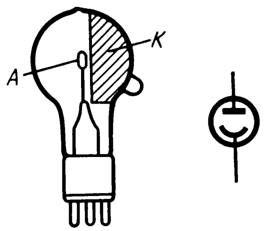


Fig. 17-1. Vacuum phototube and its diagram symbol

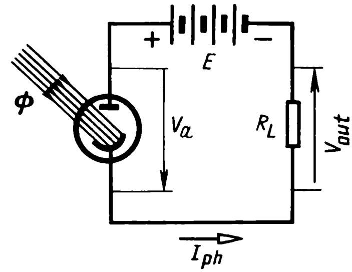


Fig. 17-2. Connection of a phototube in a circuit

light incident on a *P-N* junction causes the generation of electron-hole pairs, the junction field separates them, and an emf is produced across the *P-N* junction.

Photodiodes operate either with an external bias source (the current mode), or without an external bias because they themselves can generate an emf (the photovoltaic or voltage mode).

The most commonly used vacuum phototubes are *oxygen-cesium* and *antimony-cesium* types.

An oxygen-cesium vacuum phototube (Fig. 17-1) consists of an evacuated glass envelope whose inner surface, except a small window to pass light, is coated with a layer of silver (a base layer) and a semiconductor layer of cesium oxide which serves as the cathode, K. In antimony-cesium phototubes, the base layer of antimony is coated with a semiconductor layer.

The anode, A, is shaped into a ring because it should not prevent light from reaching the cathode.

Gas-filled phototubes are only of the oxygen-cesium type. They differ from vacuum phototubes in only that, after evacuation and prior to sealing, the envelope is filled with an amount of argon at a low pressure.

Connecting a phototube to a load resistor, R_L, and a power source (Fig. 17-2) applies an anode voltage, V_a, across the phototube, and establishes an electric field between the anode and cathode. If light is now allowed to strike the cathode through the window, the field will force the electrons to move from the cathode to the anode. In this way a *pho-*

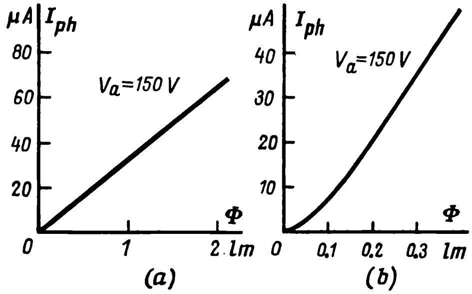


Fig. 17-3. Luminous characteristics of (*a*) a vacuum phototube and (*b*) a gas-filled phototube

tocurrent is caused to flow in the circuit so long as the cathode is illuminated. The relationship between the photocurrent I_{ph} and the light flux Φ with the supply voltage being held constant,

$$I_{ph} = f(\Phi) \quad V \text{ being constant}$$

is called the *luminous* characteristic of a phototube. For vacuum phototubes, it is linear (Fig. 17-3*a*)

$$I_{ph} = S\Phi$$

For gas-filled phototubes, it is nonlinear (Fig. 17-3*b*).

In a gas-filled phototube, electrons ionize atoms of the filling gas, so that the number of free electrons increases and so does the photocurrent. The ratio of the anode current I_a augmented by gas ionization to the initial photocurrent I_{ph} is termed the *gas amplification factor*

$$k_g = I_a / I_{ph}$$

The gas amplification factor rises with increasing gas concentration and usually ranges from 4 to 6.

One of the important parameters of a phototube is its *luminous sensitivity* defined as the ratio of the photocurrent in microamperes to the white light flux (in lumens) supplied by a standard light source

$$S = I_{ph} / \Phi$$

The luminous sensitivity of vacuum phototubes is from 20 to 120 μA/lm; that of gas phototubes ranges from 150 to 250 μA/lm.

The photoelectric emission and, as a consequence, the photocurrent of a phototube depend on the wavelength λ of the incident light. So, in addition to the luminous sensitivity, it is customary to specify the *spectral response* or *spectral sensitivity characteristic of a phototube.* The spectral response of a phototube is the ratio of its photocurrent to the incident luminous flux at a specified wavelength

$$S_\lambda = I_{ph-\lambda}/\Phi_\lambda$$

The relationship between the luminous sensitivity of a phototube and the wavelength of the incident light is called the *spectral characteristic of the phototube*

$$S_\lambda = f(\lambda), \quad \Phi_\lambda = \text{constant and } V_a = \text{constant}$$

As is seen in Fig. 17-4, an antimony-cesium phototube is most sensitive to blue-green light (it has a peak on its spectral characteristic at $\lambda = 0.4\text{-}0.5$ μm). An oxygen-cesium phototube has two peaks: at $\lambda = 0.35$ μm and $\lambda =$ $= 0.8$ μm.

The anode volt-ampere characteristics (or, simply, anode characteristics) of a vacuum phototube (Fig. 17-5), which represent the relationship between the photocurrent and anode voltage

$$I_{ph} = f(V_a), \quad \Phi \text{ being constant}$$

are nonlinear. When the voltage rises, the photocurrent increases rapidly first, then more slowly, until it ceases rising altogether (saturation is said to take place).

With gas phototubes (Fig. 17-6), an increase in the anode voltage causes the anode characteristic to bend upwards, following its flat portion; this happens because of ionization.

One of the characteristic properties of phototubes is *fatigue*, that is, an impairment in their performance with time.

Since the photocurrent produced by phototubes is low (about several milliamperes), it is customary to use them in conjunction with tube or transistor amplifiers.

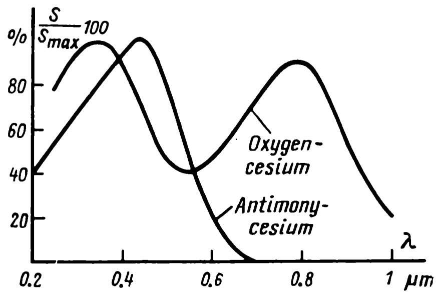


Fig. 17-4. Spectral characteristics of vacuum phototubes

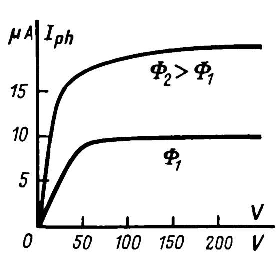


Fig. 17-5. Anode characteristics of a vacuum phototube

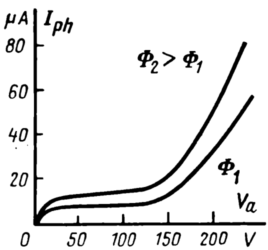


Fig. 17-6. Anode characteristics of a gas phototube

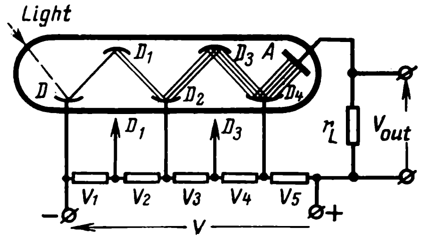
Phototubes have found wide application in various fields of electronics, automation, television, sound motion-pictures, instrumentation, etc.

Some of the simple circuits utilizing phototubes are discussed in Chapter 21.

17-2. The Photomultiplier Tube

A photomultiplier tube (or a multiplier phototube) is a photoemissive tube, that is, one utilizing the external photoeffect, whose photocurrent is multiplied by secondary emission.

A photomultiplier tube (Fig. 17-7) consists of a glass envelope containing a number of secondary-emitting electrodes (dynodes), D_1, D_2 and so on, between the photocathode



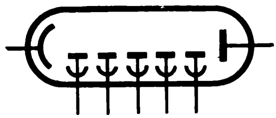


Fig. 17-7. Construction of a photomultiplier and its diagram symbol

and the output electrode, usually called the collector. The surface of the dynodes is coated with a layer of an emitting material. Every next dynode operates at a higher positive potential than the preceding one, typically about 100 V higher. The number of secondary electrons emitted by each dynode exceeds the number of primary electrons striking its surface. The ratio of the number of secondary to that of primary electrons is called the secondary-emission coefficient (production coefficient or the ratio of secondary emission) designated k_m; it usually is 3 or 4 per stage (that is, per dynode). If a multiplier phototube has n stages (or dynodes), its output current might be amplified k_m^n times. However, the maximum output current of a photomultiplier tube does not exceed a few tens of milliamperes.

The sensitivity of a photomultiplier tube may be as high as 100 A/lm. Photomultiplier tubes are used to measure faint luminous fluxes down to about 10^{-8} lm.

Soviet-made photomultiplier tubes have a varying number of stages (dynodes); they are designated ФЭУ-1 through ФЭУ-19.

17-3. Photoresistors

A photoresistor (also called a bulk photoconductor) is a semiconductor device whose resistance is appreciably decreased by incident light. This happens because the semiconductor absorbs luminous energy which ionizes its atoms and increases the number of free electrons and holes, so its resistance drops.

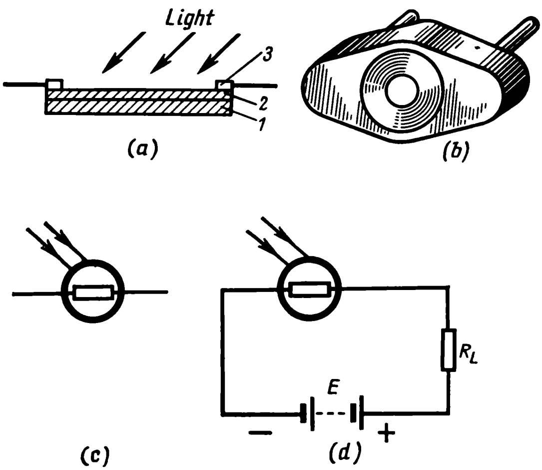


Fig. 17-8. Photoresistor
(*a*) construction; (*b*) general appearance; (*c*) diagram symbol; (*d*) circuit diagram

A photoresistor (Fig. 17-8, *a* and *b*) consists of a glass baseplate, *1*, coated (by evaporation in a vacuum) with a thin layer of a semiconductor material, *2*, and two metal electrodes, *3*, brought outside at its edges.

The semiconductor layer is given a coat of clear varnish to protect it from moisture and mechanical damage. The baseplate is enclosed in a case with two pins to which the electrodes are connected. The symbol and circuit diagram of a photoresistor are shown in Fig. 17-8*c* and *d*.

The semiconductor materials used for photoresistors are lead sulphide, cadmium selenide and cadmium sulphide. The Soviet-made photoresistors using these materials are designated as follows (Russian letters throughout); ФСА, ФСД and ФСК, respectively.

It is advisable to use the first type in the infra-red and the other types in the visible region of the spectrum.

In the absence of illumination, a low current, known as the *dark current*, flows in a photoresistor. The resistance corresponding to this current is known as the *dark resistance*; it ranges from hundreds of kilohms to several magohms. When the photoresistor is illuminated, a current, known as

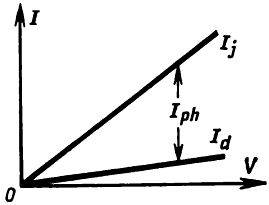


Fig. 17-9. Volt-ampere characteristic of a photoresistor

the *joint current*, flows in it. The difference between the joint and dark currents is called the photocurrent: $I_{ph} = I_j - I_d$.

Photoresistors have the same resistance in both directions and can only operate when driven by an external supply.

Photoresistors are characterized by luminous sensitivity, $S = I_{ph}/\Phi$, which is measured in μA/lm.

The luminous sensitivity is by two orders of magnitude greater than that of vacuum or gas phototubes.

The volt-ampere characteristic of a photoresistor (Fig. 17-9)

$$I_{ph} = f(V), \quad \Phi = \text{constant}$$

is usually linear.

Photoresistors have a considerable time lag, a nonlinear dependence of the photocurrent on luminous flux (the *luminous characteristic* $I_{ph} = f(\Phi)$ at $V = $ constant), and an electrical resistance strongly dependent on temperature, which is their major drawback.

Photoresistors have found wide use in industrial electronics, automation and computers.

Several simple circuits utilizing photoresistors will be discussed in Chapter 21.

17-4. Semiconductor Photovoltaic Cells

A crystal photovoltaic (or barrier-layer) cell is a semiconductor device which generates an emf when illuminated by light. This emf is known as the *photo-emf*. For its operation, a barrier-layer photocell or, which is the same, the rectifier photocell depends on the barrier layer existing between *P*- and *N*-type semiconductors.

When light hits the surface of a photovoltaic cell near the *P-N* junction, the crystal is ionized and new electron-

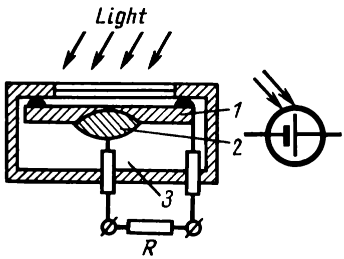


Fig. 17-10. Germanium barrier-layer photocell and its diagram symbol

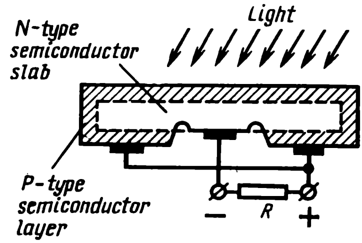


Fig. 17-11. Silicon photovoltaic cell

hole pairs are generated. The electric field of the *P-N* junction ($\mathscr{E}_j$) forces the electrons to move to the *N*-layer and the holes to the *P*-layer. This produces an excess of holes in the *P*-layer and of electrons in the *N*-layer. The resulting potential difference (photo-emf) between the *P*- and *N*-layers gives rise to a current, *I*, flowing from the *P* to the *N*-electrode in the external circuit. This current depends on the number of electrons and holes and, therefore, on the incident flux.

Figure 17-10 is a sketch of a *germanium barrier-layer photocell*. It consists of an *N*-type germanium plate, *1*, into which a dot of indium, *2*, is molten. During manufacture, a *P*-type region is formed in the germanium wafer, just above the indium do so a *P-N* junction is produced at the boundary between the *P*-type area and the germanium. The germanium layer is so thin that light easily passes through it on to the barrier layer. The photocell is built into an acrylic plastic case encapsulated in an insulating compound, *3*, through which two leads are brought outside.

A silicon photovoltaic cell (Fig. 17-11) consists of a silicon slab doped with an impurity which produces *N*-type conduction. The slab surface is doped (by diffusion in a vacuum) with boron which produces a *P*-type layer 2 μm thick. Silicon photovoltaic cells are arranged into solar batteries which convert solar energy directly into electricity. Their efficiency is about 11%. In particular, they are used in artificial satellites to supply their radio equipment.

Barrier-layer photocells have a sensitivity of up to 10 mA/lm. Their main advantage over other photocells is that

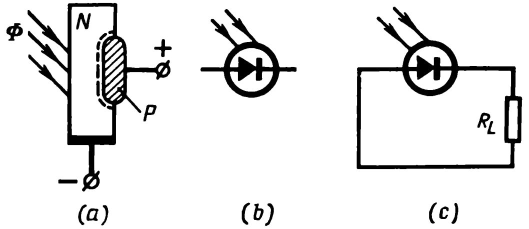


Fig. 17-12. (*a*) Photodiode; (*b*) its diagram symbol and (*c*) a circuit for operation in the voltage mode

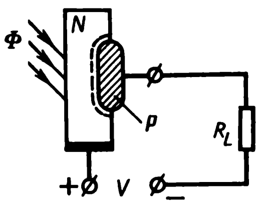


Fig. 17-13. Photodiode connected in a circuit for operation in the current mode

they do not require a power source. Barrier-layer photocells are widely used in various fields of electronics, automation, instrumentation, etc.

A photovoltaic cell having two electrodes separated by a *P-N* junction is called a *photodiode.*

As already noted in passing, photodiodes can operate in any one of two modes, namely the current mode and the photovoltaic or voltage mode. In the former case, they need and, in the latter case, they do not need an external bias source for their operation.

Figure 17-12 shows a sketch of a photodiode, its diagram symbol and circuit for operation in the photovoltaic or voltage mode.

Referring to Fig. 17-12*c*, when light hits the photodiode, additional electron-hole pairs are generated; some of them reach the *P-N* junction where the junction field drives the holes into the *P* region, whereas the electrons remain in the *N* region because they cannot overcome the potential barrier. Thus, the holes and electrons are accumulated in the *P* and *N* regions, respectively, so that a potential difference, the photo-emf, is established between the electrodes. It may

be as high as 1 V. If the circuit is connected to a load (Fig. 17-12*c*), a current will flow through it.

In the current mode, the photodiode is reverse-biased (Fig. 17-13). In the absence of illumination, a low reverse current, known as the *dark current*, flows through the diode. Light falling on the N region of the diode generates electron-hole pairs. The holes reach the P-N junction and, forced by its electric field, pass to the P region. Thus, light gives rise to a minority-carrier current from the N to the P region; this is a photocurrent. A change in the circuit current, which depends on the illumination of the diode, results in a voltage drop across the load proportional to the luminous flux incident on the photodiode. The photodiode operating in the current mode is similar to a photoresistor having a high integrated sensitivity. For example, the Soviet-made ФД-К1 silicon diode has a sensitivity of 4 to 5 mA/lm, and type ФД-2 germanium diode 20 to 25 mA/lm. The former has a dark current of 1 to 3 μA and the latter, 10 μA.

Chapter Eighteen Rectifiers

18-1. Half-Wave Rectification

Rectification is the conversion of alternating current to direct (pulsating) current. It is carried out by devices which have a very low forward and a very high reverse resistance. As already noted, such devices are called rectifiers. These may be crystal diodes, vacuum diodes, gas diodes, etc.

Figure 18-1 shows the volt-ampere characteristic of an ideal rectifier with zero forward resistance, R_f, and an infinitely high reverse resistance, R_r. The characteristic has two regions: Oa—along the positive current (y-) axis, and Ob—along the negative voltage (or x-) axis (for the reverse voltage).

Figure 18-2a gives a piecewise-linear approximation of the volt-ampere characteristic of a diode (segments bO and Oa). This characteristic may be derived from an equivalent circuit consisting of an ideal rectifier and a series resistor which represents the forward resistance, R_f, of a practical diode (Fig. 18-2b).

If we apply a sinewave voltage, $v = V_m \sin \omega t$, to an ideal rectifier ($R_f = 0$) connected in series with a load resistor, R_L (Fig. 18-3a), the current in the circuit during the positive half-cycles ($v > 0$) (Fig. 18-3b) will be

$$i = v/(R_f + R_L) = v/R_L = V_m \sin \omega t/R_L = I_m \sin \omega t \quad (18\text{-}1)$$

During the negative half-cycles ($v < 0$), there is no current flowing in the circuit because the rect ifer presents an infinitely large reverse resistance, that is

$$i = v/(R_r + R_L) = v/\infty = 0 \quad (18\text{-}1a)$$

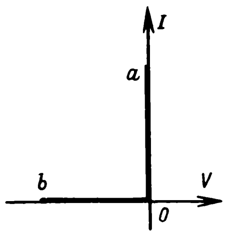


Fig. 18-1. Volt-ampere characteristic of an ideal rectifier

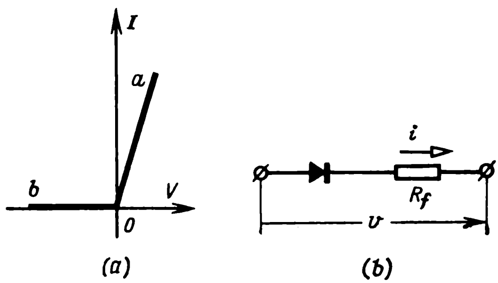


Fig. 18-2. (*a*) Volt-ampere characteristic of a rectifier and (*b*) its equivalent circuit

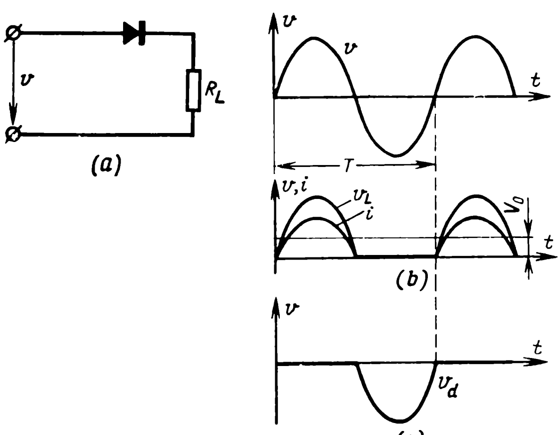


Fig. 18-3. (*a*) Rectifier connected in series with a load resistor; (*b*) waveforms for v, v_L and i; (*c*) waveform for rectifier voltage v_r

Thus, the circuit passes only one half-wave of sinusoidal current during each cycle or period (Fig. 18-3*b*).

During the positive half-cycles, the load voltage is equal to the applied voltage, $iR_L = v_L = v$, and the voltage across the rectifier is zero because its forward resistance is $R_f = 0$ (Fig. 18-3*c*). During the negative half-cycles, the load voltage is zero (Fig. 18-3*b*) because the circuit current i is zero, and the voltage across the rectifier (Fig. 18-3*c*) is equal to the applied voltage. The load voltage has a waveform similar to that of the current. As is seen, both the load voltage and the circuit current are *pulsating*.

In the analysis of electric circuits carrying nonsinusoidal, especially pulsating, currents and voltages, it is customary to use Fourier's theorem. According to this theorem, a periodical quantity may be expanded into the sum of a constant (time-invariant) component and a number of sinusoidal components differing in amplitude, frequency and epoch angle (or initial phase).

A sinusoidal component having the same frequency as the given nonsinusoidal quantity is called the *fundamental* or the *first harmonic*. The sinusoidal quantity at twice the fundamental frequency is called the *second harmonic*.

By applying Fourier's expansion half-wave rectified current may be written

$$\begin{aligned} i &= I_m/\pi + \frac{I_m}{2}\sin\omega t - \frac{2I_m}{3\pi}\cos 2\omega t - \frac{2I_m}{3\times 5\pi}\cos 4\omega t \ \ldots \\ &= I_0 + I_{1m}\sin\omega t - I_{2m}\cos 2\omega t - I_{4m}\cos 4\omega t - \ldots \end{aligned} \tag{18-2}$$

As is seen, the Fourier series contains the d.c. component

$$I_0 = I_m/\pi \tag{18-3}$$

equal to the average current $I_0 = \frac{1}{T}\int_0^T i\,dt$, the fundamental whose amplitude is $I_{1m} = I_m/2$ and whose frequency, f, is equal to that of the applied voltage, and even harmonics with frequencies $2f$, $4f$, and so on,

The rms current in the same circuit can be determined, using the definition given by Eq. (5-10)

$$I=\sqrt{\frac{1}{T}\int_0^{T/2} i^2\,dt}=\sqrt{\frac{1}{T}\int_0^{T/2} I_m^2\sin^2\omega t\,dt}$$

$$=I_m/2=\frac{\pi}{2}I_0 \tag{18-4}$$

The rms load voltage is

$$V_L = V_m/2$$

The d.c. component of the load voltage or, which is the same, the average rectified load voltage is

$$V_0=I_0R_L=I_mR_L/\pi=V_m/\pi=V\sqrt{2}/\pi=0.45V \tag{18-5}$$

Using Eq. (18-5), we can determine the rms voltage across the circuit from the given rectified voltage, V_0.

The peak-inverse voltage across the rectifier (Fig. 18-3*c*) is equal to the maximum voltage across the circuit

$$V_{p\text{-}i}=V_m \tag{18-6}$$

The total power in terms of the rms input current and rectifier voltage is

$$S=IV=\frac{\pi}{2}I_0\frac{\pi}{\sqrt{2}}V_0=\frac{\pi^2}{2\sqrt{2}}P_0$$

$$=\text{(approx.) }3.5P_0 \tag{18-7}$$

Thus, the apparent power, S (sometimes called the design secondary power of a transformer), is 3.5 times the power due to the rectified load current, $P_0 = I_0V_0$.

The ripple in a pulsating voltage or current is stated in terms of the *ripple factor*, q, defined as the ratio of the amplitude of the current or voltage fundamental to the respective d.c. component:

$$q=I_{1m}/I_0 \quad \text{or} \quad q=V_{1m}/V_0 \tag{18-8}$$

For a half-wave rectifier, the ripple factor is

$$q=I_{1m}/I_0=(I_m/2)/(I_m/\pi)=\pi/2=1.57 \tag{18-8a}$$

A major disadvantage of half-wave rectification is a large amount of ripple in the load current and voltage, that is,

an excessively large alternating component in the load current and voltage. This can be reduced by means of smoothing filters.

Example 18-1. Find the alternating voltage that must be applied to a half-wave rectifier (Fig. 18-3) to obtain a rectified voltage of 225 V.

Solution.

According to Eq. (18-5), the supply voltage is

$$V = V_0/0.45 = 225/0.45 = 500V$$

18-2. Full-Wave Rectification

Full-wave rectifier circuits are used more often than half-wave circuits because they make a better use of a power source (a transformer) and produce a lower ripple in the load current and voltage.

Figure 18-4*a* gives a full-wave rectifier circuit in which the transformer has a centre tap.

Leads *1* and *2* of the transformer secondary are connected to the anodes of two rectifiers whose cathodes are connected via a load resistor, R_L, to the centre tap on the same winding.

During the first half-cycle, the potential at point *1* is higher than that at the centre tap, O, and the current is flowing through the first diode and lead.

During the second half-cycle, the potential at point *2* exceeds that at the centre tap, O, and the current is flowing via the second rectifier and load. Thus, two half-waves of current pass through the load in the same direction during each cycle (Fig. 18-4*c*). Therefore, the direct component of the load current is twice as great as in the case of a half-wave rectifier circuit (18-3), that is

$$I_0 = 2I_m/\pi \tag{18-9}$$

and the rms load current is

$$I = \sqrt{\frac{2}{T}\int^{T/2} i^2\,dt} = I_m/\sqrt{2}$$

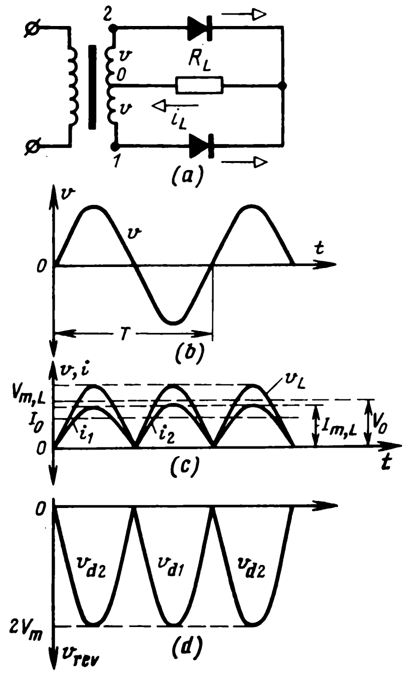


Fig. 18-4. (*a*) Full-wave rectifier circuit; (*b*) supply voltage waveform; (*c*) load voltage and current waveforms; (*d*) reverse voltage waveform

Since the current is flowing through each rectifier only during a half-cycle, the average current of the rectifier is one half the average load current.

The direct component of the load voltage is

$$V_0 = I_0 R_L = 2I_m R_L/\pi = \frac{2}{\pi} V_m$$

$$= \frac{2\sqrt{2}}{\pi} V = 0.9V \qquad (18\text{-}10)$$

or twice that in the case of half-wave rectification [see Eq. (18-5)].

During the first half-cycle, the first diode is conducting. Since its forward resistance is zero, $R_f = 0$, the load voltage is equal to the voltage across half of the transformer secondary, $v_L = v$. During the same half-cycle, a negative voltage equal to the sum of the voltage across one half of the

transformer secondary and the load voltage, $v_{d2} = v + v_L = 2v$, is applied to the second (nonconducting) diode.

Thus, a peak-inverse voltage of a full-wave rectifier is twice the amplitude of the voltage across the winding

$$V_{p\text{-}i} = 2V_m = 2\sqrt{2}V = 2\sqrt{2}\,\frac{\pi}{2\sqrt{2}}\,V_0 = \pi V_0 \qquad (18\text{-}11)$$

The active power in the load is

$$P_a = I^2 R_L = I_m^2 R_L/2 = V_m I_m/2 = VI = S \qquad (18\text{-}12)$$

As is seen, with full-wave rectification, the active power is equal to the apparent power.

The power in the load due to the rectified current is

$$P_0 = I_0 V_0 = (2I_m/\pi)(2V_m/\pi) = (2\sqrt{2}/\pi)^2 IV$$
$$= 0.81\,S$$

The ripple factor for any rectifier, except the half-wave type, is given by

$$q = 2/(m^2 - 1) \qquad (18\text{-}13)$$

where m is the number of phases in the rectifier.

A full-wave rectifier is treated as a two-phase circuit in which the phase difference is a half-cycle ($T/2$).

Applying Eq. (18-13) to a full-wave rectifier, we get

$$q = 2/(2^2 - 1) = 2/3 = 0.667$$

Major drawbacks of this circuit are a high peak-inverse voltage and poor utilization of the transformer secondary, because current is flowing in each half of the winding only during one half-cycle.

Full-wave rectifiers are used to supply tubes in radio receivers, TV sets, amplifiers and low-power oscillators.

Figure 18-5 shows a bridge-type full-wave rectifier.

Each arm of the bridge contains a rectifier diode. The sinusoidal voltage, $v = V_m \sin \omega t$, to be rectified is applied across one pair of bridge junctions, a and c, while the load resistor, R_L, is connected across the other pair of the bridge junctions, b and d. During the positive half-cycles, when the potential at point a exceeds that at point c ($v > 0$), diodes D_1 and D_2 are conducting, their resistance being $R_{D1} = R_{D2} = 0$, and diodes D_3 and D_4 are turned off because

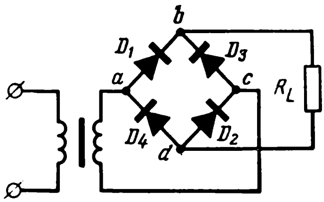


Fig. 18-5. Bridge-type full-wave rectifier circuit

a negative voltage equal to the supply voltage v is applied to each of them. Thus, during the positive half-cycles, the current

$$i = v/(2R_f + R_L) = v/R_L = V_m \sin \omega t/R_L = I_m \sin \omega t$$

from the source is flowing through D_1, R_L and D_2. During the negative half-cycles ($v < 0$), diodes D_3 and D_4 are conducting, whereas diodes D_1 and D_2 are not. Now the current is flowing from the source via D_3, R_L and D_4.

Thus, two unidirectional half-waves of current are flowing via the load during each cycle, and the load voltage is the sum of two half-waves of the same sign. Therefore, as in the previous case, the direct component of the load current according to Eq. (18-9) is

$$I_0 = 2I_m/\pi$$

and the direct component of the load voltage is

$$V_0 = I_0 R_L = 2V_m/\pi = 0.9\,V \qquad (18\text{-}14)$$

In contrast to the previous case, the peak-inverse voltage across the rectifier is equal to the maximum input voltage

$$V_{p\text{-}i} = V_m \qquad (18\text{-}15)$$

As in the previous case, the average current in each diode is one-half of the average load current.

A bridge rectifier offers a number of advantages over the previous rectifier circuit: the peak-inverse voltage across the diodes is halved; the circuit makes a better use of the transformer, because the secondary current is flowing during the whole cycle; and, finally, the circuit can operate without a transformer.

A major disadvantage is that it requires four rectifiers, whereas the former circuit uses only two.

The bridge-type rectifier circuit is usually based on crystal diodes.

Example 18-2. Given. A bridge rectifier circuit built around diodes having a peak-inverse voltage of 350 V.

To find. The rms supply voltage, V, and the rectified voltage, V_0.

Solution.

According to Eq. (18-15), $V_{p\text{-}i} = V_m$, so the rms supply voltage is

$$V = V_m/\sqrt{2} = 350/1.41 = \text{(approx.) } 248 \text{ V}$$

Recalling Eq. (18-14), we may write for the rectified voltage

$$V_0 = 0.9\ V = 0.9 \times 248 = 223 \text{ V}$$

18-3. Three-Phase Rectifiers

Figure 18-6*a* shows a likely arrangement for a three-phase rectifier.

The starts A, B and C of the secondary windings of a three-phase transformer are connected to the anodes of three diodes whose cathode leads meet a common junction, O'. The load resistor, R_L, is connected between the two common junctions, O and O'.

Through each diode, current is flowing one-third of a cycle when the voltage of the transformer phase containing this diode exceeds those of the other two phases. In Fig. 18-6*b*, this voltage is represented by a heavy curve made up of the tops of the sinusoidal phase voltages. So long as one rectifier is conducting, the other two are turned off and their resistances are infinity. So the current path is from the transformer phase, via the conducting rectifier to the load. With an ideal rectifier and resistive load, the circuit current is $i = v/R_L$, and the load voltage is equal to the phase voltage, $v = v_L = iR_L$. So the plot given in Fig. 18-6*b* also represents the load voltage, v_L. The same plot, but drawn to some other scale, would hold for the load current.

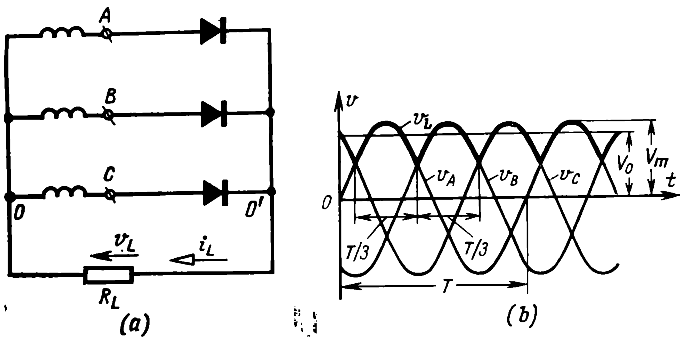


Fig. 18-6. (*a*) Circuit diagram of a three-phase rectifier and *b*) phase voltage waveforms

Thus, in contrast to a full-wave rectifier where the load voltage and current vary from zero to peak values, a three-phase rectifier produces a substantially smaller ripple in the load current and voltage.

In order to determine the average rectified (load) voltage which is given by the average ordinate of the v-curve, we find the area bounded by this curve, the ordinates corresponding to times $T/12$ and $5T/12$, and the x-axis, and divide it by the base, that is, $T/3$,

$$V_0 = \frac{1}{T/3} \int_{T/12}^{5T/12} V_m \sin \omega t \, dt = \frac{3\sqrt{3}}{2\pi} V_m$$

$$= 0.827\, V_m = 1.17\, V \tag{18-16}$$

The average rectified load current is

$$I_0 = V_0/R_L = (3\sqrt{3}/2\pi)(V_m/R_L) = (3/\pi)(\sqrt{3}/2)\, I_m$$
$$= 0.827\, I_m \tag{18-17}$$

Since the current is flowing through each rectifier diode only during one-third of a cycle, its average value is one-third the load current

$$I_r = I_0/3 \tag{18-18}$$

The maximum secondary currents of the transformer or the peak rectifier currents equal to them can be derived

from Eq. (18-17):

$$I_m = I_{r,\,m} = 2\pi I_0/3\sqrt{3} = 1.21\,I_0 \qquad (18\text{-}19)$$

or, recalling Eq. (18-18), we finally obtain

$$I_m = I_{r,\,m} = 1.21\,I_0 = 3.63\,I_r \qquad (18\text{-}19a)$$

The peak-inverse voltage is equal to the peak line voltage

$$V_{p\text{-}i} = \sqrt{3}\,V_m = \sqrt{3}\,(2\pi V_0/3\sqrt{3}) = 2\pi V_0/3$$
$$= 2.09\,V_0$$

The ripple factor in this case is

$$q = 2/(3^2 - 1) = 0.25$$

18-4. Selection of Diodes for Rectifier Circuits

In most cases, we select and design semiconductor rectifier circuits on the basis of the following specified quantities: primary alternating voltage V, average rectified (load) voltage V_0, and average rectified load current I_0.

After a circuit configuration has been selected, we determine the circuit in each rectifier, the type of the rectifier and the number of parallel branches. Then we find the peak-inverse rectifier voltage V_{p-i} and the number of diodes connected in series. Finally, we determine the input voltage V_2 and the turns ratio of a transformer. When selecting rectifier diodes, it is important not to exceed the maximum safe values of V_{p-i}, $I_{f\text{-max}}$ and I_0 given by the manufacturers in their data sheets.

Let us consider several examples.

Example 18-3. Given. The limiting current of a Soviet-made type Д226 diode, $I_{d,\text{lim}} = 200$ mA; the required rectified current, $I_r = 900$ mA.

To find. The number of type Д226 crystal diodes necessary to produce the specified rectified current in a full-wave rectifier.

Solution.

Since $1/2\; I_r > I_{d,\,\text{lim}}$, several diodes must be connected in parallel. The number of diodes connected in parallel is given by

$$m = 1/2\; I_r/k_i I_{d,\,\text{lim}}$$

where $k_i = 0.5$ to 0.8 is the load current factor; in our case, we adopt $k_i = 0.75$.

Thus, the number of parallel branches is $m = 450/0.75 \times \times 200 = 3$.

As the diodes somewhat differ in forward resistance, series resistors are placed in tandem with them to equalize their currents (see Fig. 16-17). The values of the series resistors are given by

$$R_s \geqslant \frac{V_{f,\,av}(m-1)}{mI_{d,\,\lim} - 1.1I_r} = \frac{1\,(3-1)}{3\times 300\times 10^{-3} - 1.1\times 450\times 10^{-3}}$$

$$= \text{(approx.) } 4 \text{ ohms}$$

where $V_{f,\,av}$ is the average forward voltage drop across a diode.

Example 18-4. Given. $V_{a.c} = 700$ V, $V_{p-i} = 300$ V, $I_{r\text{-max}} = 300$ μA.

To find. The number of Soviet-made type Д226Б crystal diodes in a single-phase bridge rectifier.

Solution.

The peak-inverse sinusoidal voltage is

$$V_m = \sqrt{2}\,V_{a.c} = \sqrt{2} \times 700 = \text{(approx.) } 1000 \text{ V}$$

Since $V_m > V_{1p-i}$, the required number of diodes connected in series (Fig. 16-18) is found to be $n = V_m/k_v V_{1p-i} = 1000/0.7 \times 300 =$ (approx.)5. Here, $k_v = 0.5$ to 0.8 is the load voltage factor. In our case, $k_v = 0.7$.

In order to equalize the reverse resistances of the diodes, they are shunted by resistors. Their resistance is given by

$$R_{sh} \leqslant \frac{nV_{1p\text{-}i} - 1.1V_m}{(n-1)\,I_{r\text{-max}}} = \frac{5\times 300 - 1.1\times 1000}{(5-1)\times 300\times 10^{-6}} = 300 \text{ kilohms}$$

18-5. Response of an RC Network

(a) Forced Response of an RC Network

Let there be a network consisting of an unchanged capacitor of capacitance C, and a resistor of resistance R, connected to a d.c. source of voltage V (Fig. 18-7).

It is assumed that prior to applying power to the RC network, the capacitor is uncharged, the voltage across it is

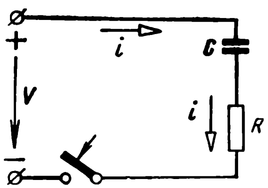


Fig. 18-7. Explaining the forced response of an RC network

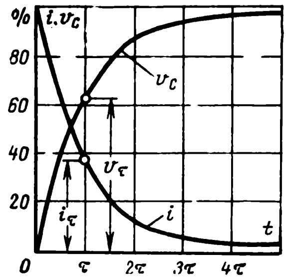


Fig. 18-8. Current and voltage waveforms in the forced response of an RC network

$v_C = 0$. So at the initial time ($t = 0$), the voltage drop across the resistor R is equal to V, and a current

$$i = V/R = I$$

is flowing in the network.

After power is applied, a flow of current, i, is started accompanied by a gradual build-up of charge, Q, on the capacitor. As a result, a voltage, $v_C = Q/C$, is developed across it, bringing down the voltage drop across the resistor R

$$iR = V - v_C$$

in accordance with Kirchhoff's second (voltage) law. Hence the current

$$i = (V - v_C)/R$$

will go down, and the rate of charge build-up will decrease, too, because

$$i = dQ/dt \tag{18-20}$$

With time, the capacitor continues charging, but its charge Q and voltage v_C increase at a slower rate (Fig. 18-8), so the circuit current gradually decreases in proportion to the potential difference, $V - v_C$.

A sufficiently long (theoretically, infinitely long) time later, the capacitor voltage equals the source voltage, the current falls to zero, and the capacitor is fully charged.

In practice, charging is considered to be complete when the current has decreased to one per cent of its initial value,

V/R, or, which is the same, when the capacitor voltage has reached 99% of the source voltage, V.

The charging time increases with increasing circuit resistance R as it limits the current, and with increasing capacitance C, because a greater capacitance accumulates a larger charge. The rate of charging is characterized by the *time constant of the network*

$$\tau = RC \tag{18-21}$$

As τ increases, the rate of charging slows down.

The time constant has the dimension of time

$$[\tau] = [RC] = \Omega \times \mathrm{C/V} = \Omega \times \mathrm{A} \times \mathrm{s/V} = \mathrm{s}$$

In τ seconds after power is applied, the capacitor voltage reaches about 63% of its full value, and in 5τ the capacitor is considered to be fully charged.

In charging, the capacitor voltage is

$$v_C = V - V \exp(-t/\tau) = V[1 - \exp(-t/\tau)] \tag{18-22}$$

or, in words, it is equal to the difference between the d.c. source (or steady-state) voltage and what is known as the *force-free* or *transient voltage*, $V \exp(-t/\tau)$ which decreases with time exponentially from V to 0 (Fig. 18-8).

The charging current of the capacitor is

$$i_C = (V/R) \exp(-t/\tau) = I \exp(-t/\tau) \tag{18-23}$$

The capacitor current, i_C, too, decreases exponentially with time from its initial value $I = V/R$ (Fig. 18-8).

(b) Free Response of an RC Network

Now assume that prior to the switching the capacitor C was allowed to charge to the source voltage V, and trace the response of the network as the capacitor discharges through the resistor R (Fig. 18-9). To initiate a discharge, the switch is moved from position *1* to position *2*. Initially, the circuit current is $i = V/R = I$, the capacitor begins to discharge and its voltage begins to go down. As v_C decreases, the circuit current, $i = v_C/R$, also decreases (Fig. 18-10). After $5\tau = 5RC$, the capacitor voltage and the circuit cur-

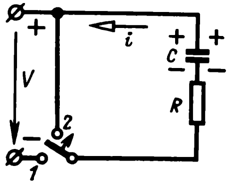


Fig. 18-9. Explaining the free response of an RC network

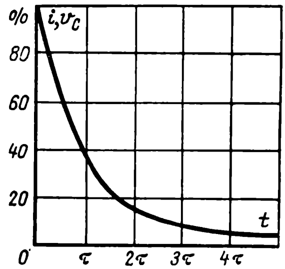


Fig. 18-10. Current and voltage waveforms in the free response of an RC network

rent will have decreased to one per cent of their initial values, and the capacitor is considered to be fully discharged.

The voltage across the capacitor during discharge has a zero steady-state component, and its force-free or transient voltage is

$$v_C = V \exp(-t/\tau) \tag{18-24}$$

that is, it decreases exponentially with time (Fig. 18-10).

The same reasoning applies to the discharge current of the capacitor

$$i_C = -v_C/R = -I \exp(-t/\tau)$$

that is, it also decreases with time according to the exponential law (Fig. 18-10).

All of the energy stored by the capacitor as an electric field on charge is dissipated in the resistor R as heat during discharge.

The electric field of a charged capacitor disconnected from its power source cannot remain the same for a long time because the dielectric and insulation of the capacitor have a certain conductivity.

The discharge of a capacitor due to the conduction through its dielectric and insulation is known as *self-discharge.* The self-discharge time constant τ of a capacitor does not depend on the shape of and spacing between the capacitor plates.

Thus, we have seen that both the forced and force-free response of an RC network is accompanied by *transient processes* (or simply, transients).

18-6. Wave Rectifiers

A rectified voltage (or current) consists of a direct and an alternating component.

In most cases, d.c. loads utilize only the direct voltage (or current) component. The alternating component is not used; indeed, it causes considerable power losses, reduces the efficiency of devices, and is sometimes accompanied by other undesirable effects. This is why every effort is made to reduce the alternating component, more commonly known as the ripple. The ripple is reduced by smoothing or ripple filters connected between the output of a rectifier and its load.

The efficiency of a ripple filter is defined as the ratio of its input to output ripple factors

$$S = q_{in}/q_{out} \tag{18-25}$$

and is called the *smoothing factor*; it shows how effective a filter is in reducing the ripple.

The most commonly used types of ripple filter are *capacitor-input*, *choke-input*, *LC* and *RC filters.*

A capacitor-input filter is essentially a capacitor of capacitance C connected in parallel with the load resistor, R_L (Fig. 18-11*a*).

The rectifier voltage v_r is equal to the difference between the source voltage v and the capacitor voltage v_C, that is, $v_r = v - v_C$. The current can flow through the rectifier only when $v - v_C > 0$. Therefore, at time t' (Fig. 18-11*b*) when $v - v_C > 0$, the capacitor begins charging, and the charge current i_C and load current i_L begin flowing through the rectifier, so that $i_r = i_C + i_L$.

At time t'', when $v - v_C = 0$, the capacitor ceases charging, the voltage v falls to below v_C, and the capacitor begins to discharge through the load resistor R_L. In this case, the voltage across the capacitor decreases exponentially with time

$$v_C = V_{C0} \exp(-t/\tau)$$

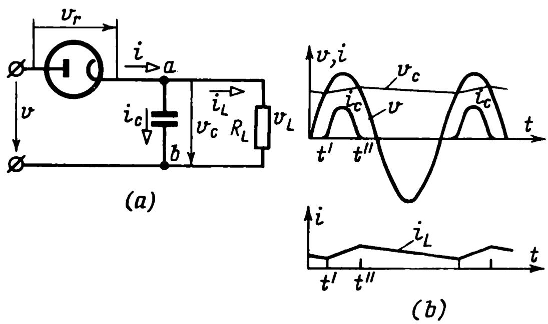


Fig. 18-11. (*a*) Capacitor-input filter in a rectifier circuit and (*b*) voltage and current waveforms

where V_{C0} = voltage across the capacitor at time t″ when the rectifier is turned off

$\tau = CR_L$ = discharge time constant of the circuit, or the time during which v_C decreases to $1/e = 1/2.72$ of its original value

When $\tau \gg T$, the voltage decreases insignificantly until the rectifier is driven to conduction again. During the same time, the discharge current of the capacitor which is also the load current ($i_C = i_L$) varies as little. Then the chain of events repeats itself all over again. In this way, the load voltage $v_L = v_C$ and the load current $i_L = v_L/R_L$ are smoothed.

During the negative half-cycles, the source voltage is combined with the load voltage, so that the peak-inverse voltage of the diode may be

$$V_{p\text{-}i} = 2V_m \tag{18-26}$$

Capacitor-input filters are used in low-power rectifiers.

As its name implies, a choke-input filter (Fig. 18-12) consists of a choke (which is an inductor coil wound on an iron core) having inductance L, inductive reactance $x_L = \omega L$, and resistance R_f, which is connected in series with a load resistor R_L.

The choke-input filter operates effectively in circuits handling high currents, provided that $m\omega L \gg R_L$ and $R_f \ll$ $\ll R_L$. Then the direct component at the input differs but

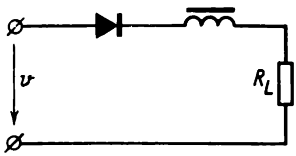


Fig. 18-12. Choke-input filter in a rectifier circuit

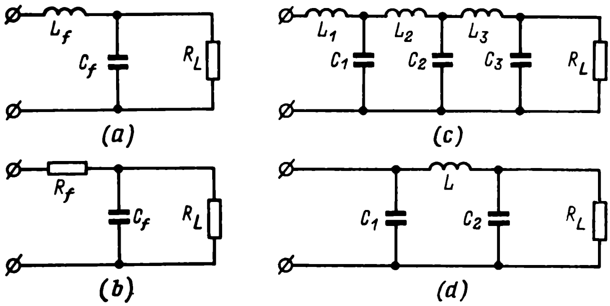


Fig. 18-13. Filters

(*a*) L-section *LC*-filter; (*b*) L-section *RC*-filter; (*c*) three L-section *LC*-filter; (*d*) pi-section *LC*-filter

little from the direct component at the filter output, because

$$V_{0,\,in}/V_{0,\,out} = I_0\,(R_f + R_L)/I_0 R_L = \text{(approx.)}\ R_L/R_L = 1$$

More effective smoothing can be achieved with L-section filters (Fig. 18-13*a*) consisting each of a choke, L_f, connected in series with a shunt arm containing a load resistor, R_L, and a capacitor, C_f. Better smoothing is achieved because the choke eliminates a good proportion of the ripple, so that the a.c. component across the shunt arm is greatly reduced. The capacitor connected in parallel with the load resistor offers a much lower resistance to the alternating current component than R_L, so this component, that is, the ripple is considerably reduced in the load.

In some cases, the L-section *LC*-filter can be replaced by an L-section *RC*-filter (Fig. 18-13*b*). However, the smoothing factor of this filter is much lower because the ripple is smoothed less due to the absence of a choke, and the direct component of the load voltage is decreased by the voltage drop across the filter resistor, R_f. In order to

reduce this voltage drop, the value of R_f is taken to be 20% of R_L.

On the other hand, the use of a resistor instead of a choke reduces the mass, size and cost of a filter.

Major disadvantages of the RC-filter are voltage and power losses in the filter resistor. So they are used in low-power rectifiers where these losses can be neglected.

If the smoothing factor of a single L-section filter is not sufficient, use is made of filters consisting of two or three L-sections (Fig. 18-13*c*). The resultant smoothing factor is the product of the smoothing factors of all filter sections

$$S = S_1 S_2 S_3 \tag{18-27}$$

Apart from L-section filters, vacuum-tube and semiconductor rectifiers use pi- (Π- or π-) section filters (these filters are not used in gas-filled rectifiers).

A pi-section filter (Fig. 18-13*d*) is a combination of a capacitor-input filter and an L-section LC-filter.

The smoothing factor of a pi-section filter exceeds that of an L-section filter. It is the product of the smoothing factors of the capacitor-input filter (S_{C1}) and the L-section filter (S_L)

$$S = S_{C1} S_L \tag{18-28}$$

18-7. Thyristor Rectifiers

Widely used rectifier systems (such as used for electric traction or d.c. electric power drives) should not only convert a single- or three-phase current into a direct current, but also ensure a continuous regulation of the average rectified voltage and current.

Until recently, use for this purpose was mainly made of control-grid mercury-pool rectifiers. At present, they have been superseded by controlled semiconductor rectifiers which are smaller, more efficient and have a better performance in comparison. They make it possible to regulate output voltage over a wide range at low power in control circuits.

In Sec. 18-2, we have examined the design (Fig. 18-4) and operation of a full-wave rectifier with a centre-tapped transformer. Figure 18-14 shows the circuit of a similar rectifier in which the diodes are replaced by thyristors.

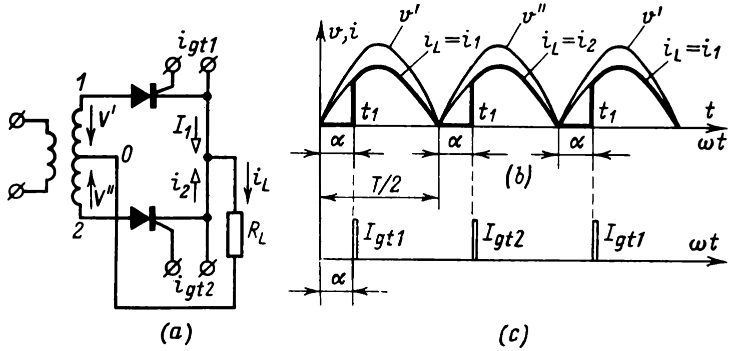


Fig. 18-14. (*a*) Circuit diagram of a full-wave rectifier based on thyristors; (*b*) waveforms for V', V'', I and load current; (*c*) waveforms for trigger pulses

In the circuit built around diodes, the current i_1 is flowing through the first rectifier during the positive half-cycles of the voltage v' impressed by the first half of the transformer secondary. The current i_2 is flowing in the second rectifier also during the positive half-cycles of the voltage v'' supplied by the second half of the secondary, which is shifted in phase by a half-cycle relative to the voltage v'. Thus, two positive half-waves of current are flowing through the load during each cycle of the alternating voltage.

In the circuit based on thyristors, a rectifier is turned on just as a control (or trigger) pulse is applied.

If the trigger pulses are applied at the beginning of a cycle (the firing angle $\alpha = 0$), the current will flow in the circuit in much the same manner as in the diode circuit. If trigger pulses generated by an automatic control unit lag behind the beginning of a cycle by $\alpha > 0$, the conduction of a thyristor will be delayed for the corresponding part of the cycle, $t_1 = \alpha_1/\omega$, and the duration of current flow during each half-cycle will be reduced by the same part of the cycle (Fig. 18-14*b*). Thus, as α is increased, the time $(T/2 - \alpha_1/\omega)$ during which the current is flowing through the rectifiers (the conduction angle) decreases, and the average load current and voltage decrease, too, because $v_L = i_L R_L$.

Figure 18-15*a* gives the circuit of a simple controlled three-phase rectifier built around thyristors. It differs from

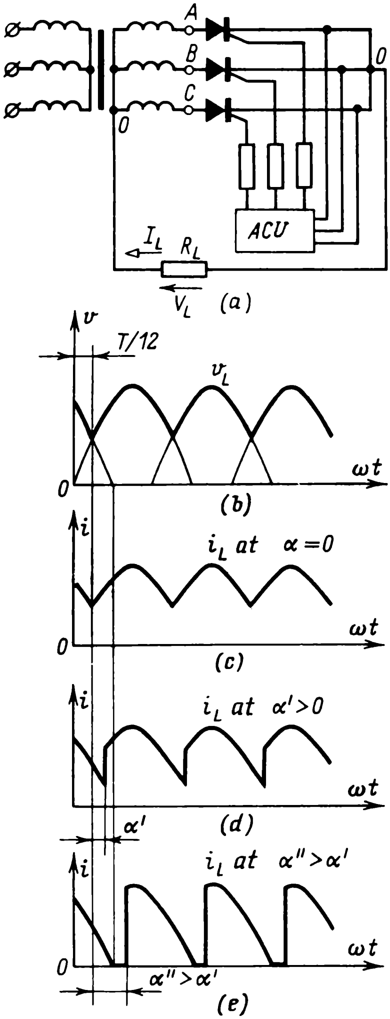


Fig. 18-15. (*a*) Circuit diagram of a simple three-phase thyristor rectifier; (*b*) load voltage waveform; (*c*) load current waveform at $\alpha = 0$; (*d*) load current waveform at $\alpha > 0$; (*e*) load current waveform at $\alpha'' > \alpha'$

the previous circuit (see Fig. 18-6) in that it uses thyristors instead of diodes. At the respective instants of a cycle, an automatic control unit applies trigger pulses to the gates of the thyristors. The instant at which each thyristor is triggered, its *ON*-time during each cycle, and, as a consequence, the average rectified current and voltage are all determined by the firing angle α. The firing angle α is reckoned from that instant of a cycle when the voltage across the phase containing the thyristor being turned on reaches the value of the leading phase voltage. In our case, this instant is one-twelfth of a cycle from its beginning.

When $\alpha = 0$, the current and voltage waveforms for each thyristor will, for one-third of a cycle, be the same as they are in a rectifier based on diodes (Fig. 18-15*b* and *c*). At low firing angles α, the current will be flowing continuously in the load (Fig. 18-15*d*). The instant at which the current flow is transferred from one thyristor to another is determined by the firing angle α. At large firing angles, the flow of load current will be intermittent (Fig. 18-15*e*).

As the firing angle is increased, the average current in and average voltage across the load go down.

Chapter Nineteen Audio-Frequency Amplifiers

19-1. General

An amplifier is an electronic device intended to boost weak input signals in voltage, current and/or power by drawing power from an external supply sources. This process can be controlled by a vacuum tube or a transistor.

Amplifier circuits are many and diverse and may be classified in several ways. For example, (1) by the frequency range, they can be classed as audio-frequency (a.f.), radio-frequency (r.f.) or direct-current (d.c.) amplifiers (d.c. amplifiers operate in the frequency range from zero to several hertz); (2) by the number of stages, they can be classed into one-stage, two-stage and multi-stage amplifiers; (3) by the quantity being amplified, they can be classed into voltage, power and current amplifiers.

In a voltage amplifier, the output signal gains power mainly due to voltage amplification. In power and current amplifiers, the output signal gains power mainly due to current amplification. Power amplifiers are usually the output or final stages of multistage amplifiers.

According to the manner in which the stages are coupled, amplifiers can be classified under three types, namely direct-coupled, resistance-capacitance (or RC-) coupled, and transformer-coupled.

Audio-frequency amplifiers are the most widely used in industrial electronics. Their main parameters are: gain factor, bandwidth, power output or output voltage, and efficiency.

The gain factor (or, simply gain) of a voltage amplifier is the ratio of the output to the input voltage

$$k_v = V_{out}/V_{in} \qquad (19\text{-}1)$$

The gain of a multistage amplifier is equal to the product of the gains of the individual stages

$$k_1k_2 \dots k_n = (V_1/V_{in})(V_2/V_1) \dots (V_{n-1}/V_{n-2})(V_{out}/V_{n-1}) = k_v \tag{19-2}$$

The gain of a power amplifier is the ratio of the output to the input power

$$k_p = P_{out}/P_{in} \tag{19-3}$$

In multistage amplifiers, the gain may be 10^6 and higher.

The loudness of audio sounds perceived by man varies in proportion to the logarithm of the corresponding variation in sound energy. That is why the gain is often expressed in logarithmic units known as bells (B). A gain of 1 B is a power ratio of 10/1, for which the common (or decimal) logarithm is unity. So, power variations in bells may be written

$$S_p\,(\mathrm{B}) = \log_{10}(P_{out}/P_{in}) \tag{19-4}$$

More often, use is made of the decibell which is one-tenth of a bell; so

$$S_p\,(\mathrm{dB}) = 10\log_{10}(P_{out}/P_{in}) \tag{19-4a}$$

The gain (in dB) of a multistage amplifier is the sum of the gains of its stages

$$S = S_1 + S_2 + \dots + S_n \tag{19-5}$$

Power gain in decibells can be found, if we recall that $P = V^2/R = I^2R$:

$$\left.\begin{aligned} S_V &= 10\log_{10}(V_{out}/V_{in}) \\ S_I &= 10\log_{10}(I_{out}/I_{in}) \\ S_p &= 20\log_{10}(V_{out}/V_{in}) \\ &= 20\log_{10}(I_{out}/I_{in}) \end{aligned}\right\} \tag{19-6}$$

Example 19-1. Given. The input power of a three-stage amplifier is $P_{in} = 0.01$ W, the output power is $P_{out} =$ $= 100$ W.

To find. The gain in dB.

Solution.

The power gain (as ratio) is

$$k_p = P_{out}/P_{in} = 100/0.01 = 10{,}000$$

The gain (in dB) is

$$S_p = 10 \log_{10}(P_{out}/P_{in}) = 10 \log_{10}(100/0.01) = 40 \text{ dB}$$

The bandwidth of an amplifier is the frequency range within which variations in the gain do not exceed specified limits. For a.f. amplifiers, the bandwidth is from several hertz to a few tens of kilohertz.

The output power of an amplifier is the power which it delivers to a load

$$P_{out} = V^2_{out\text{-}m}/2R_L \tag{19-7}$$

The nominal or *rated output power of an amplifier* is the maximum power delivered to a load, at which distortion does not exceed specified limits.

The electrical efficiency of an amplifier is the ratio of its output power P_{out} to the power P_0 expended by the plate supply source of the amplifier tube

$$\eta_{el} = P_{out}/P_0 \tag{19-8}$$

The overall efficiency of an amplifier is the ratio of its output power to the total power applied

$$\eta_{ov} = P_{out}/P_{tot} \tag{19-9}$$

The electrical efficiency of an a.f. power amplifier ranges from 40 to 70%. The overall efficiency is considerably lower than the electrical efficiency.

In practical amplifiers, an electrical signal is distorted as it passes through. The distortion may be with respect to frequency, amplitude (in which case, it is called nonlinear distortion), or phase.

Frequency distortion is the distortion in signal waveform which results when frequencies in a complex wave are not amplified or attenuated by the same amount. As is seen from the frequency response curve of an a.f. amplifier (Fig. 19-1), the midband gain (k_m) is constant, but it rolls off as the frequency rises to f_h (the upper limit frequency) and decreases to f_l (the lower limit frequency). Frequency distortion is stated in terms of the ratio of k_m to the gain at a given frequency

$$M = k_m/k_l \tag{19-10}$$

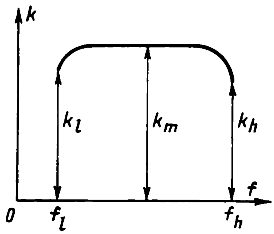


Fig. 19-1. Frequency response of an amplifier

This ratio shows how much the gain at a given frequency departs from the midband gain. The allowable amount of frequency distortion is usually 1.25.

Amplitude or *nonlinear distortion* is the distortion in signal waveform that occurs when an amplifier puts out frequencies that are not present in the input signal. This distortion is due to the nonlinear characteristics of the vacuum tubes or transistors in amplifiers or of their loads.

Nonlinear distortion is characterized by the ratio between the square root of the sum of the squared harmonics in the load voltage or current of an amplifier to the fundamental of its load voltage or current

$$\nu = \sqrt{V_2^2 + V_3^2 + V_4^2 + \ldots + V_n^2}/V_1 \qquad (19\text{-}11)$$

Depending on the purpose of an amplifier, nonlinear distortion may vary from 0.05 to 15%.

Phase distortion is the distortion that occurs when the phase of the output signal differs from that of the input signal. Phase distortion is due to inductive and capacitive elements (L and C) used in an amplifier.

A. TRANSISTOR AMPLIFIERS

19-2. Practical Common-Emitter Amplifiers

Power for the input and output circuits of a transistor can be supplied by either two independent sources (see Fig. 16-25*b*) or one source (Fig. 19-2).

For normal operation, a transistor needs a direct voltage applied between its emitter and base. It is termed the *base bias voltage*, and it varies from transistor to transistor, the most commonly used value being 0.5 V. Using the input

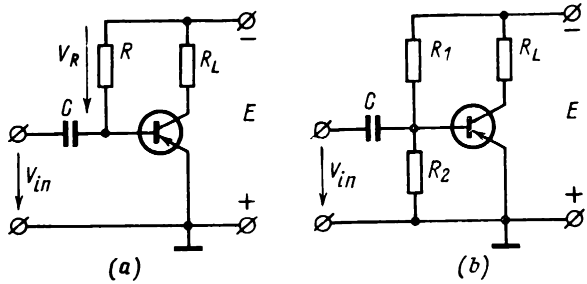


Fig. 19-2. Base bias supply circuits

characteristic of a transistor (Fig. 16-27), we can find the respective base current, I_{bA}, of the amplifier. In the common-emitter circuit (Fig. 19-2*a*), the bias voltage is taken from a resistor, R, connected between the base and the negative side of the power source. In the no-signal condition, the sum of the voltage drop across the resistor, $I_{bA}R$, due to the constant base current I_{bA}, and the base-to-emitter voltage $V_{be,A}$ is equal to the source voltage E

$$I_{bA}R + V_{be,A} = E \tag{19-12}$$

Hence, the necessary resistance is

$$R = (E - V_{be,A})/I_{bA}$$

or, recalling that $V_{be,A} \ll E$, we finally get

$$R = (\text{approx.})\ E/I_{bA}$$

The circuit shown in Fig. 19-2*b* uses a voltage divider, R_1R_2, to obtain the base bias voltage. The value of the first resistor is

$$R_1 = (E - V_{be})/(I_d + I_{bA}) = (\text{approx.})\ E/(I_d + I_{bA})$$

and that of the second is

$$R_2 = V_{be,A}/I_d$$

where I_d is the current in the voltage divider.

As is stated in Sec. 16-4, the parameters and operation of transistors are markedly affected by temperature variations. In particular, an increase in temperature brings about an increase in currents, which, in turn, leads to a change in the

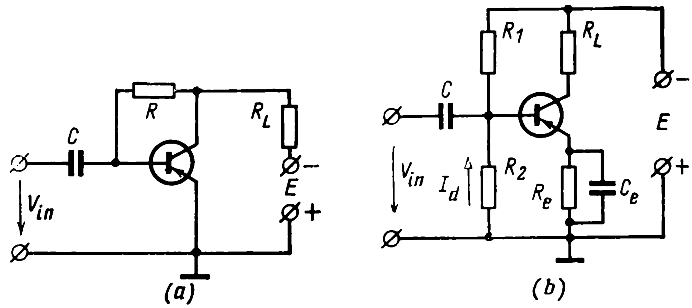


Fig. 19-3. Collector temperature compensators

operating conditions of transistors. To avoid this, use is made of temperature compensation.

Figure 19-3*a* shows what has come to be known as the collector compensator. In this circuit, the resistor, R, supplying the bias voltage (Fig. 19-2*a*) is placed in the collector lead. Should a rise in temperature increase the collector current I_c, the voltage drop, I_cR_L, across the load resistor R_L will grow, thereby causing V_{ce} and V_{be} to decrease. Accordingly, the collector current I_c and base current will go down practically to the initial value. In this way, the rise in I_c caused by an increase in temperature is compensated for.

Figure 19-3*b* gives another, more advanced, form of compensator known as the emitter compensator. Here, compensation is effected by a resistor in the emitter lead, R_e. The voltage drop across this resistor, $V_e = I_{eA}R_e$, and the voltage, $V_2 = I_dR_2$, due to the flow of the divider current, I_d, across the resistor R_2 are directed in opposition to each other, so the base bias voltage is their difference, $V_{be,A} = V_2 - V_e$. In this way, R_e provides negative direct-current feedback. Owing to this feedback, the decrease in currents caused by an increase in temperature causes $V_e = I_{eA}R_e$ to build up, thereby bringing down the base bias voltage, and the currents are forced to decrease.

The capacitor C_e shunting the emitter resistor R_e eliminates a.c. feedback. The capacitive reactance offered by the capacitor at any operating frequency must be consider-

ably lower than the resistance of the emitter resistance, $x_C \ll R_e$.

19-3. The Quiescent (Q) Point. Current and Voltage Waveforms

Let us consider the operation of the loaded common-emitter circuit (Fig. 19-4), because this is the most commonly used configuration for an amplifier stage.

The input (or signal) voltage V_{in} is applied between the base and emitter. The power source, E_c, which is utilized to boost the signal in power is placed in the output circuit. The resistor R_b in the base circuit sets the base current which positions the operating (quiescent or Q) point on the input characteristic (Q' in Fig. 19-5*b*) and on the load line (Q in Fig. 19-5*a*) plotted on the output characteristics of the transistor.

In the no-signal condition ($V_{in} = 0$), a quiescent current, I_{bA}, is flowing in the circuit and gives rise to a direct collector current, I_{cA}, whose value is determined by the point where the load line intersects the output characteristics corresponding to the base current, $I_{bA} = 40\ \mu A$.

Figure 19-5*a* shows a family of output characteristics for a transistor and its load line, MN, which represents the relationship between the collector current, I_c, and the collector-emitter voltage, V_{ce}, with the load resistance, R_L, held constant.

In order to construct the load line, it is necessary to know the source voltage E_c and the load resistance R_L. The load line is constructed as the line joining points lying on the axes of coordinates.

From Sec. 16-10, we know that the source voltage E_c is split between the collector junction resistance r_c and the load resistor R_L, so $E_c = V_{ce} + I_cR_L$. From this output circuit equation, it follows that $E_c = V_{ce}$ at $I_c = 0$, because $I_cR_L = 0$. Laying off $V_c = E_c$ along the x-axis, we get one terminal point (N) of the load line. Assuming $V_{ce} = 0$ we obtain $E_c = I_cR_L$, whence $I_c = E_c/R_L$. Laying off this value along the y-axis, we get the other terminal point (M) of the load line. The line joining the points M and N

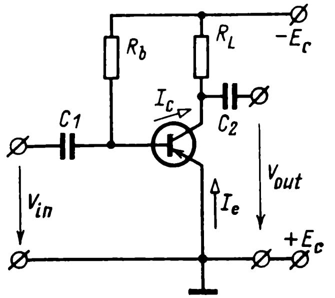


Fig. 19-4. Circuit diagram of a common-emitter voltage amplifier

is the load line. The region between points B and C is called the operating region; it corresponds to the maximum possible output current I_c, output voltage V_{ce} and output power P_{out}. A projection of the operating region BC on the y-axis is the peak-to-peak amplitude of the collector current $(2I_{c,m})$, and that on the x-axis is the peak-to-peak amplitude of the collector-emitter current $(2V_{ce,m})$.

Assuming that the input and output quantities are sinusoidal, we locate the quiescent (Q) point in the middle of the region BC. As its name implies, the Q point determines the quiescent collector current, I_{cA}, and the quiescent collector-emitter voltage, $V_{ce,A}$.

The input characteristic of a transistor relates the base current to the base-emitter voltage, $I_b = f(V_{be})$, with V_{ce} held constant. In each particular case, one ought to take the input characteristic that corresponds to the selected collector-emitter voltage. On the other hand, input characteristics plotted at various values of $V_{ce} > 0$ differ very little; also, it is usual for data sheets to give input characteristics for only one value of this voltage, say, $V_{ce} = 5V$. So, in approximate calculations, only one input characteristic taken from a data sheet or handbook is used.

If we transpose the points Q, B and C from the output to the input characteristic, we obtain the respective points Q', B' and C' (Fig. 19-5,b). Projections of the regions $Q'C'$ and $Q'B'$ into the axes of coordinates give the amplitudes of the input voltage, V_{be}, and input current, I_b, respectively.

The current gain is $k_I = I_c/I_b$, the voltage gain $k_V = V_{ce}/V_{be}$, and the power gain $k_P = k_I k_V$.

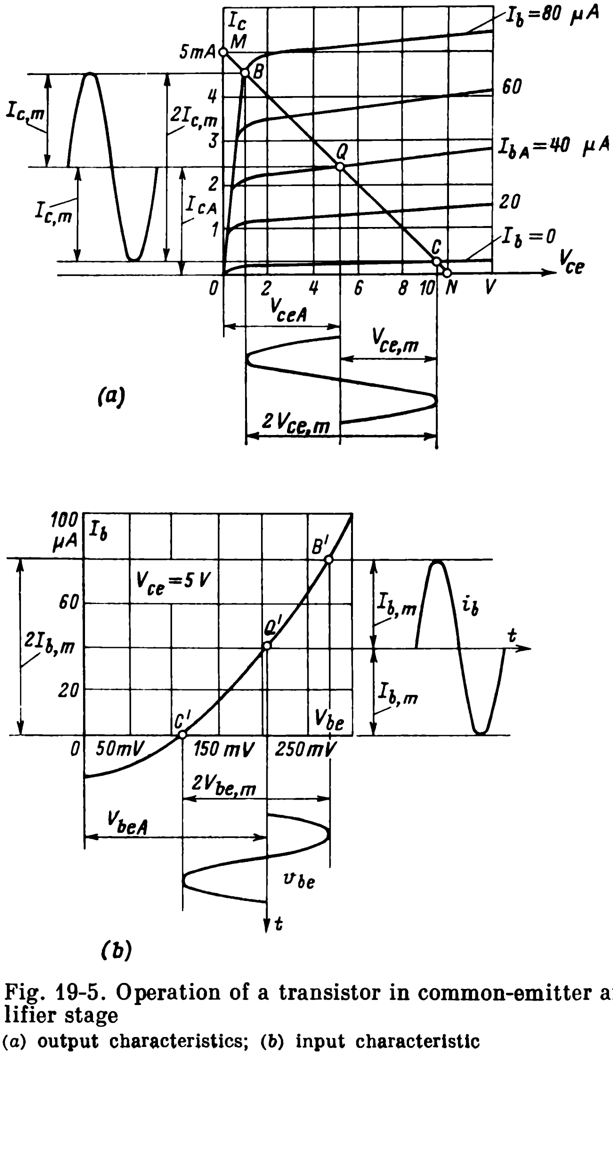


Fig. 19-5. Operation of a transistor in common-emitter amplifier stage

(*a*) output characteristics; (*b*) input characteristic

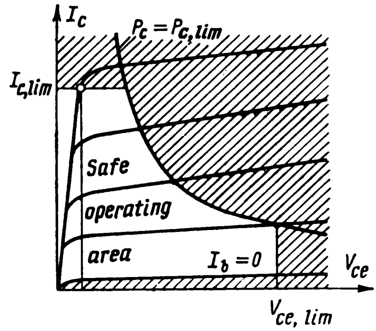


Fig. 19-6. Safe operating area

For proper use of transistors, it is important to know the limits of safe operating conditions. These limits bound the safe operating area on the transistor characteristics. Figure 19-6 shows this area for a transistor connected in a common-emitter circuit. It is bounded by the limiting collector current $I_{c,\,lim}$, limiting collector-emitter voltage $V_{ce,\,lim}$, limiting power $P_{c,lim}$, and base current $I_b = 0$. Besides, if nonlinear distortion is to be low, the operating area should not extend into nonlinear regions of the characteristics.

Depending on the initial operating conditions and the base current amplitude, the collector current can flow either during some part or the entire cycle of input voltage. Accordingly, a transistor amplifier can operate in four classes called Class A, Class B, Class AB, and Class C.

Class A. The Q point is positioned in the middle of the load line BC (Fig. 19-5*a*). The amplitude of the input current I_{infm} should be less than the transistor quiescent current I_{c0}.

In Class A operation, distortion is minimal but the efficiency does not exceed 40%. This class of operation is utilized in all voltage amplifier stages and low-power output stages.

Class B. The Q point is positioned in an area where the collector current is close to I_{c0} (Fig. 19-7), that is, near the x-axis. The quiescent output current is nearly zero. The transistor conducts during a half-cycle or (in Russian usage) with current cut-off which is characterized by the cut-off

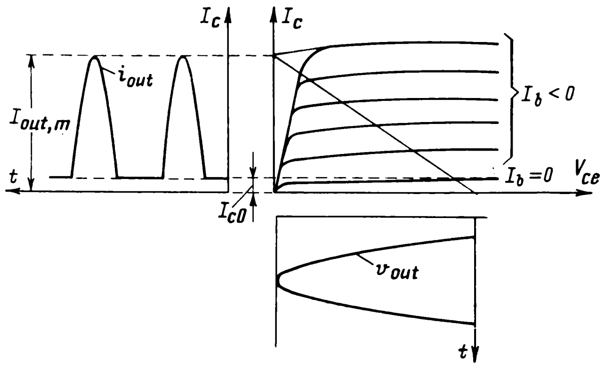


Fig. 19-7. Class B operation

angle θ equal to half the conduction angle*. In Class B, the cut-off angle is 90° and the conduction angle is 180°. In Class B operation, nonlinear distortion is great and the efficiency may be up to 70%. This class of operation is utilized in the push-pull type of amplifier where current flows alternately in either of the transistors.

Class AB is intermediate between Class A and Class B. The Q point is positioned at the bottom of the load line, so the output signal is greatly distorted. This class of operation is likewise used mainly in push-pull amplifiers.

19-4. Frequency Response of Amplifiers

The alpha (emitter-to-collector) current gain of a transistor remains practically unvaried over a wide frequency range, but the amplifying properties of the transistor decline with increasing frequency. Figure 19-8 shows the relationship between the alpha current gain and frequency. The frequency at which the alpha current decreases to $1/\sqrt{2} = 0.707$ of its low-frequency value is called the *alpha cutoff frequency*, f_α.

One of the main reasons for the decrease in alpha gain with increasing frequency is that the transit time of mino-

* In the US and UK literature, only the conduction angle is considered.— *Translator's note.*

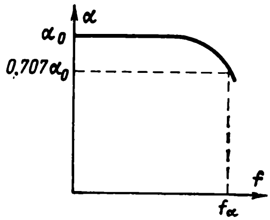


Fig. 19-8. Relationship between alpha current gain and frequency

rity carriers through the base is finite. Injected all at the same time, they reach the collector junction at different times, so the waveform of the output signal is distorted and its amplitude decreases. Also, this causes a phase delay between the emitter and collector currents.

The second reason for the decrease in alpha gain is the effect of the collector junction capacitance C_c which may be considered to be connected in parallel with the collector-junction resistance, r_c. At low frequencies, the capacitive reactance of C_c is high and its effect may be ignored; at high frequencies, the capacitive reactance $1/\omega C_c$ goes down, and its shunting effect brings down the alpha current gain.

19-5. Multistage Transistor Amplifiers

When the specified gain cannot be achieved by a single amplifier stage, use is made of multistage amplifiers.

Figure 19-9 shows the circuit of a widely used RC-coupled two-stage transistor amplifier.

The circuits and component functions of the first and second amplifier stages were discussed in Sec. 19-2.

In the first stage, C_b is a blocking capacitor intended to separate the d.c. component due to the base bias voltage (supplied by the power source) from the signal (a.c. component). The Q point in this stage is positioned by resistor R_{b1} and stabilized by the network $R_{e1}C_{e1}$ which applies negative d.c. feedback. R_{c1} is the load resistor in the collector circuit. C_{c1} couples the first and second stages.

The second stage differs from the first in that the Q point is positioned by a voltage divider consisting of two resistors, R_2' and R_2''. Capacitor C prevents the d.c. component

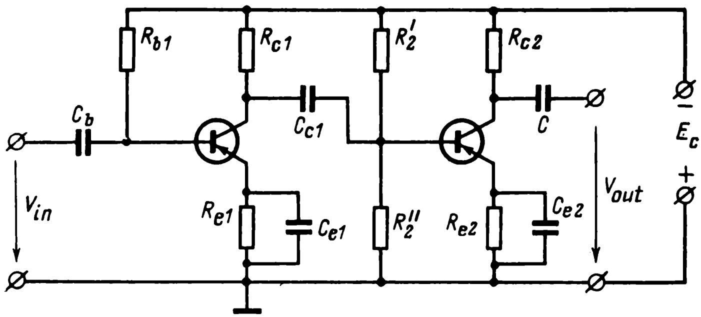


Fig. 19-9. RC-coupled two-stage transistor amplifier

of the collector current from reaching the load at the output of the second stage.

One common-emitter amplifier stage has a voltage and current gain of 10 to 20, and a power gain of 100 to 300.

As already noted, in addition to RC-coupled amplifiers use is also made of transformer-coupled amplifiers such as shown in Fig. 19-10.

The Q point is positioned by a voltage divider consisting of resistors R_1 and R_2 in the first stage, and of resistors R_1' and R_2', in the second. Both stages use emitter temperature compensation networks C_eR_e and $C_e'R_e'$ (see Sec. 19-2).

The collector circuit of the first stage contains the primary winding of a coupling transformer, Tr_1. The secondary winding of this transformer is connected via capacitor C_{c1} between the base and emitter of the second stage transistor.

The primary winding of a second transformer, Tr_2, is connected in a similar manner, and its secondary is connected to the load resistor R_L (or it may be connected between the base and emitter of a third stage).

In a common-emitter RC-coupled amplifier, the output impedance of the transistor is high (tens of kilohms) and the input impedance is low (hundreds of ohms). Accordingly, the previous stage delivers low power to the next. On the other hand, maximum power transfer requires that the output impedance of a previous stage must be equal to the input impedance of the next.

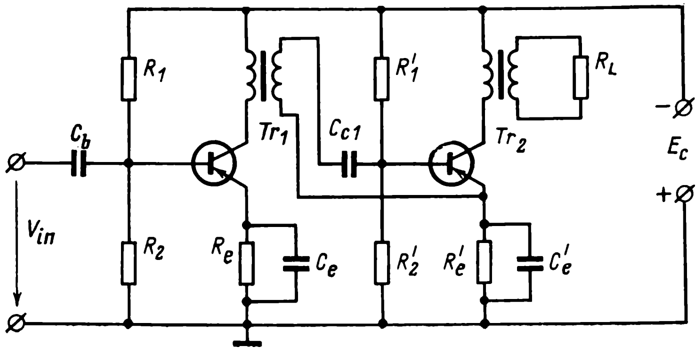


Fig. 19-10. Transformer-coupled two-stage transistor amplifier

With transformer coupling, the output impedance of each previous stage can conveniently be made equal to the input impedance of the next by an appropriate choice of turns ratio for the transformer.

Another advantage of this form of coupling is that amplifiers can use low-voltage supply sources. This is because the voltage drop across the primary winding of the transformer is considerably less than that across the load resistors, $R_L = R_c$.

The drawbacks of transformer-coupled amplifiers are: high frequency distortion, complicated design (because a transformer is more complicated than a resistor), larger size and mass, and higher cost.

19-6. The Final Transistor Amplifier Stage

An a.f. power amplifier is usually the final stage of an amplifier or a radio-receiver.

It can be loaded into an electromagnetic relay, loudspeaker, earphones, electric motor, or any other final control element.

In circuit configuration, the voltage amplifier and the power amplifier are similar, but they markedly differ in the requirements they should meet in the circuit parameters and elements they use.

A power amplifier should develop a maximum power

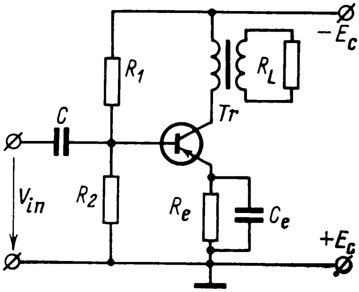


Fig. 19-11. Single-ended transistor power amplifier

(current) in a given load, and nonlinear distortion should not exceed specified limits.

Transistor power amplifiers may be of any of two types: single-ended (Fig. 19-11) or push-pull (see Sec. 19-10).

A single-ended amplifier is used when the power output must be 3 to 5 W. It operates in Class A and is connected in a common-emitter circuit which has a higher gain. In the circuit shown in Fig. 19-11, the Q point is positioned by resistor voltage divider, R_1/R_2. The effect of temperature variations is taken care of by an emitter compensation network, R_e/C_e. For maximum power transfer to the load, the circuit uses a matching transformer.

It is an easy matter to prove that maximum power transfer to the load (R_L) can be achieved when the output resistance of the signal source, R_s, is matched to the load resistance, $R_s = R_L$. Accordingly, one calls it a matched load. As amplifier characteristics are linear only within a limited region, R_L is often chosen to be 0.1 to 2 times the source resistance. In such cases, the load may be connected directly in the output (collector) circuit. More often, however, the load resistance is low, $R_L \ll R_s$. This is where one needs a matching transformer. Here, matching is achieved by adjusting the turns ratio of the transformer to suit a particular load resistance, so that power transfer to load is a maximum, and nonlinear distortion is kept within the specified limits.

If the voltage (turns) ratio of the transformer is $V_1/V_2 = w_1/w_2 = k_t$, the current ratio is $I_1/I_2 = w_2/w_1 = 1/k_t$ and the load resistance referred to the primary side is

$$R'_L = V_1/I_1 = V_2k_t/(I_2/k_t) = (V_2/I_2)\,k_t^2 = R_Lk_t^2 \quad (19\text{-}13)$$

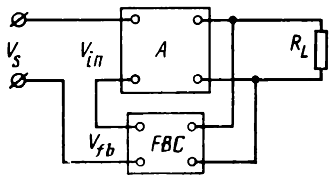


Fig. 19-12. Block diagram of a feedback amplifier

When R_L is specified in advance, the necessary turns ratio is

$$k_t = \sqrt{(0.1 \text{ to } 2)\, R_s/R_L} \qquad (19\text{-}14)$$

Example 19-1. Given. Signal source (transistor) resistance $R_s = R_c = 5000$ ohms, load resistance $R_L = 25$ ohms.
To find. The turns ratio for the matching transformer.
Solution.
The turns ratio of the transformer is

$$k_t' = \sqrt{R_s/R_L} = \sqrt{5000/25} = 14$$

As already noted in Sec. 19-1 Eq. (19-2), the overall gain factor (as ratio) of a multistage amplifier having a final stage is equal to the product of the gain factors of all amplifier stages.

The overall gain factor (in dB) of a multistage amplifier is the sum of the gain factors of all amplifier stages [see Eq. (19-5)].

19-7. Feedback Amplifiers

Feedback in amplifiers refers to the effect that the output circuit of an amplifier has on its input circuit. The circuit connecting the amplifier output to its input is called the feedback circuit. Feedback is positive if it increases the overall gain, otherwise, it is negative.

In feedback amplifiers (Fig. 19-12), the feedback voltage V_{fb} which is a portion of output voltage, V_{out}, is applied to the input.

The ratio

$$\beta = V_{fb}/V_{out}$$

is called the *feedback factor*.

The input voltage of a feedback amplifier is the sum of the signal voltage and feedback voltage

$$V_{in} = V_s + V_{fb}$$

Recalling that the gain of an amplifier having no feedback is

$$k = V_{out}/V_{in}$$

and that of a feedback amplifier is

$$k_{fb} = V_{out}/V_s$$

we finally get

$$V_{in} = V_s + V_{fb} = V_{out}/k_{fb} + V_{out}\beta = V_{out}\,(1/k_{fb} + \beta)$$

Hence, the gain factor of a feedback amplifier is

$$k_{fb} = k/(1 - \beta k) \qquad (19\text{-}15)$$

At $\beta k = 1$, positive feedback is called critical, because the gain of the amplifier becomes infinity, $k_{fb} = \infty$, and the amplifier jumps into oscillations, that is, begins to generate output voltage even in the absence of input voltage.

Negative feedback improves the stability and performance of the amplifier. In particular, it reduces nonlinear distortion. This is because a fraction of any harmonic appearing in the output signal of the amplifier and distorting the signal is applied over the feedback circuit to the amplifier input in antiphase and minimizes signal distortion.

B. VACUUM-TUBE AMPLIFIERS

19-8. The Basic Vacuum-Triode A.F. Amplifier Stage

(a) Amplification

When a triode is used in an amplifier, its plate circuit contains a plate load resistor, R_L, so any change in grid voltage causes a change in the plate load voltage.

Figure 19-13*a* shows the schematic circuit diagram of a basic amplifier stage using a vacuum triode. The plate circuit contains the plate load resistor, R_L, and a plate battery, E_p. The grid circuit contains a resistor R_g, an external source V_s of the signals to be amplified, and a grid-bias battery E_g.

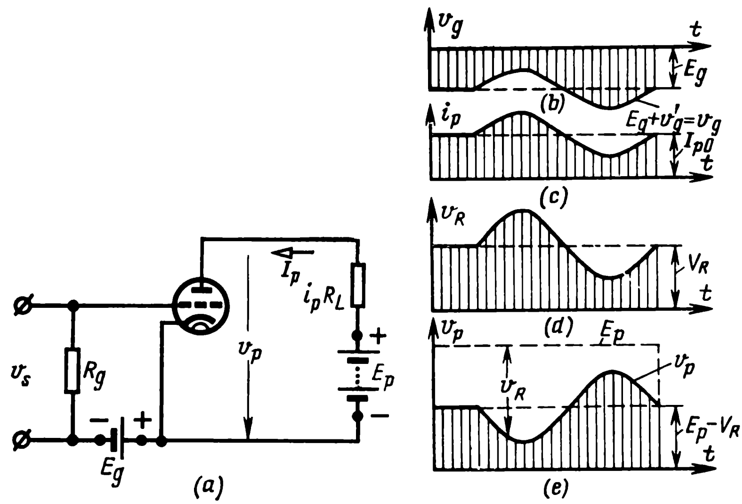


Fig. 19-13. Basic amplifier and its current and voltage waveforms
(*a*) circuit diagram; (*b*) grid voltage waveform; (*c*) plate current waveform; (*d*) load voltage waveform; (*e*) plate voltage waveform

The grid resistor R_g connects the grid to the negative side of the grid battery E_g, so the grid is at a negative potential. The signal voltage V_s is applied from the source across the grid resistor R_g. The value of R_g should be many times the internal resistance of the signal source, R_s, ($R_g \gg R_s$) so as to ensure a low voltage drop across the source. The input resistance of the circuit is stable enough when R_g is about 0.1 to 1 megohm. The grid resistor is sometimes called the grid-leak resistor because it provides a d.c. path to ground for the electrons accumulating on the grid. In the absence of R_g, the electrons would accumulate on the grid, and their charge would cut off the tube.

Let us see how the input a.c. voltage or, simply, the signal, with instantaneous values v_s is amplified. To begin with, we turn on the supply voltage E_p and adjust the negative grid supply voltage E_g (Fig. 19-13*b*), so that the grid draws no current. In the no-signal condition, a quiescent current, I_{p0} (Fig. 19-13*c*), flows in the plate circuit. The value of this current depends on the value of R_L, tube charac-

teristics, and the values of E_p and E_g. When the signal, $v_s = v'_g$, is applied to the input (Fig. 19-13*b*), the resultant grid voltage v_g consists of two components, v'_g and E_g. During the positive half-cycles, the resultant negative voltage reaches its minimum (in absolute terms), $V_{g,\,m} + (-E_g) = V_{g,\,\min}$; during the negative half-cycles it reaches its negative maximum, $-V_{g,\,m} + (-E_g) = -V_{g,\,\max}$.

Even small variations in the resultant grid voltage v_g will cause considerable variations in the plate current, so the latter increases when the negative grid voltage goes down during the positive half-cycles, and decreases when the negative grid voltage builds up during the negative half-cycles (Fig. 19-13*c*).

Thus, the direct plate current is turned into a pulsating one under the action of a small alternating grid voltage, v_g, and contains a considerable alternating component.

The plate current flowing through R_L produces a voltage drop across it (Fig. 19-13*d*)

$$v_R = i_p R_L = I_{p0} R_L + i'_p R_L = V_R + v'_R \qquad (19\text{-}16)$$

The plate voltage, v_p, is equal to the difference between the E_p, the plate supply voltage, and v_R, the voltage drop across R_L

$$v_p = E_p - v_R = E_p - (V_R + v'_R) = (E_p - V_R) - v'_R \quad (19\text{-}17)$$

Waveforms for E_p, $(E_p - V_R)$ and v_p are shown in Fig. 19-13*e*.

The a.c. component of the plate voltage $(-v'_p)$ is the output voltage (v_{out}) which is the amplified grid (v'_g) or signal (v_s) voltage. With no distortion, the output voltage waveform is an exact replica of the signal waveform. Thus, amplification consists in that a low-power alternating current flowing in the grid circuit produces high-power oscillations in the plate circuit.

It is to be noted that the plate current i_p and the plate load voltage v'_R vary in phase with the grid voltage v'_g, and the plate voltage v_p or, which is the same, the output voltage v_{out} vary in antiphase with it (Fig. 19-13). Thus, an RC-coupled amplifier operating into a resistive load inverts the phase of the input voltage (shifts it through π or 180°).

(b) Characteristics and Parameters of the Amplifier Stage

For the plate circuit shown in Fig. 19-13*a*, we may write

$$E_p - V_p = I_p R_L$$

whence the plate current is

$$I_p = (E_p - V_p)/R_L = (E_p/R_L) - (V_p/R_L) \quad (19\text{-}18)$$

Graphically, the relationship between the plate current and plate voltage is shown by a plot known as the *load line.*

To construct the load line, we need two terminal points which are located as follows. From Eq. (19-18) it follows that when $I_p = 0$ (the tube is turned off by a negative grid voltage), the plate voltage is equal to the plate supply voltage, $V_p = E_p$, because $I_p R_L = 0$. Laying off $V_p = E_p$ along the x-axis, we locate one terminal point (B in Fig. 19-14). On setting the plate current in Eq. (19-18) equal to zero, we obtain $I_p = E_p/R_L$. Laying off this value along the y-axis, we locate the second terminal point (C). Now we join the two terminal points B and C and obtain the load line. Analytically, it is described as $I_p = f(V_p)$, with E_p and R_L held constant.

The curve relating the plate current to the grid voltage for a constant plate load, R_L, with E_p held constant is termed the *dynamic transfer characteristic of the stage.* Analytically, it is described as $I_p = f(V_g)$, with E_p and R_L held constant.

Figure 19-15 shows the dynamic transfer characteristic (*abcd*) and grid-plate transfer (or static-transfer) characteristics of a triode.

When the grid voltage is equal to the cut-off voltage, $E_{g0} = V_{g,\,cut\text{-}off}$, the plate current is zero ($I_p = 0$). Naturally, the voltage drop across the load is zero, too ($I_p R_L = 0$), and the plate voltage is equal to the plate supply voltage, $V_p = E_p$. Point *a* (Fig. 19-15) is the common origin for both the static-transfer and dynamic transfer characteristics. When the grid voltage increases, a current appears in the plate circuit. If there were no plate load resistor, R_L, the plate current would follow the grid-plate transfer characteristic plotted at $V_p = E_p$. In the presence of the plate load resistor, a voltage drop, $I'_p R_L$, is developed across

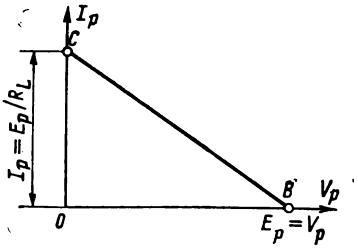


Fig. 19-14. Load line

it and the triode voltage goes down to $V_p' = E_p - I_p' R_L$. Consequently, for I_p' there should be a particular ordinate (point b) on the dynamic- and static-transfer characteristics plotted at V_p'. In a similar manner, each time the grid voltage increases, the plate current and voltage drop across the load resistor will increase, as well. As a result, the triode plate voltage V_p will consecutively go down to $V_p'' = E_p - I_p'' R_L$, $V_p''' = E_p - I_p''' R_L$, etc. When the plate voltage is V_p'', the current I_p'' will be represented by point c belonging to both the dynamic- and static-transfer characteristics; at V_p''', the current I_p''' will be represented by point d, and so on.

Thus, when the negative grid voltage decreases, the plate voltage decreases, too, and the plate current follows the dynamic transfer characteristic $abcd$ having a smaller slope than the static-transfer characteristics. The dynamic mutual

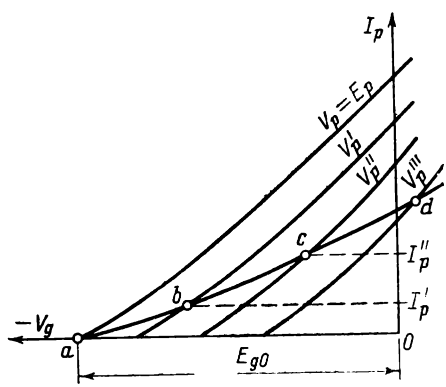


Fig. 19-15. Dynamic transfer characteristic

Example 19-2. Given. An amplifier (Fig. 19-13) built around a triode whose mutual conductance is $g_m = 4.4$ mA/V and amplification $\mu = 76$, operating into a plate load of $R_L = 100$ kilohms.

To find. The stage gain.

Solution.

According to Eq. (14-7), the a.c. plate resistance of the triode is

$$R_p = \mu/g_m = 76/4.4 = 17.3 \text{ kilohms}$$

From Eq. (19-19*a*), it follows that

$$k = \mu R_L/(R_L + R_p) = 76 \times 100/(100 + 17.3)$$
$$= \text{(approx.) } 65$$

(c) Negative Grid Biasing

In Sec. 19-8, negative grid bias was supplied by a separate source, the bias battery. However, amplifiers utilizing such a source have a greater size and mass and are more expensive. So, more often use is made of what is known as *automatic* (or *self-*) bias (Fig. 19-18) which is developed by the flow of the direct component of plate current, I_{p0}, through a resistor, R_k, in the cathode lead as a voltage drop, $V_{g,m} = V_k = -I_{p0}R_k$. As a result, the grid potential relative to the cathode automatically decreases by V_k (hence the name "automatic bias").

The a.c. component of the plate current has its path completed by a bypass capacitor, C_k, connected in parallel with the cathode resistor, R_k. The capacitive reactance of

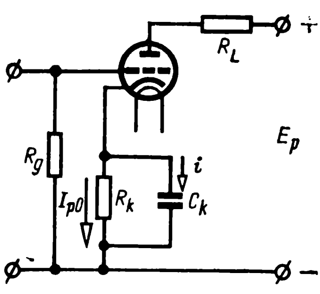


Fig. 19-18. Automatic bias

C_k must be low as compared with R_k, so that the voltage drop across R_k due to the a.c. component of the plate current would be low, and would not effect the grid bias voltage.

Usually, because the grid bias V_b is about several volts and R_k ranges from 50 to 1500 ohms, the capacitance should be

$$C_k \geqslant 20/(R_k \omega_l) \qquad (19\text{-}21)$$

where ω_l is the lowest frequency of the input signal.

(d) Classes of Amplifiers

In accordance with the position of the Q point on the grid-plate characteristic, and also depending on the a.c. input-voltage amplitude and the grid bias, V_{g0}, an amplifier may operate in Class A, Class B or Class C.

Class A (Fig. 19-19*a*). The Q point is in the middle of the linear region of the grid-plate characteristic. The signal voltage amplitude does not swing beyond the linear part of the characteristic and does not enter the positive grid voltage area because $V'_{g0} > V_{s,m}$. In Class A operation, the d.c. component of the plate current, I_{pQ}, is rather high and the plate current flows constantly. Nonlinear distortion, is low, but the amplifier efficiency does not exceed 30%. This is because the quiescent current, I_{p0}, is always higher than the peak a.c. component of the plate current, $I_{p,m}$, which determines the output power P_{out}. Class A operation is widely utilized in voltage amplifiers and single-ended power amplifiers.

Class B (Fig. 19-19*b*). The Q point is positioned at the cut-off point on the grid-plate characteristic ($V''_{g0} = V_{cut\text{-}off}$). The plate current flows only for half of each cycle. The product of the angular frequency and time t' during which the plate current varies from its peak value to zero is called the cut-off angle θ*. In Class B operation, the cutoff angle is $\theta = 90°$. In the no-signal state, the tube draws no or very small plate current, I_{p0}.

* In the US literature, the operation of a vacuum tube is specified in terms of the conduction angle (the angle of anode current flow in the UK), which is twice the cut-off angle.— *Translator's note.*

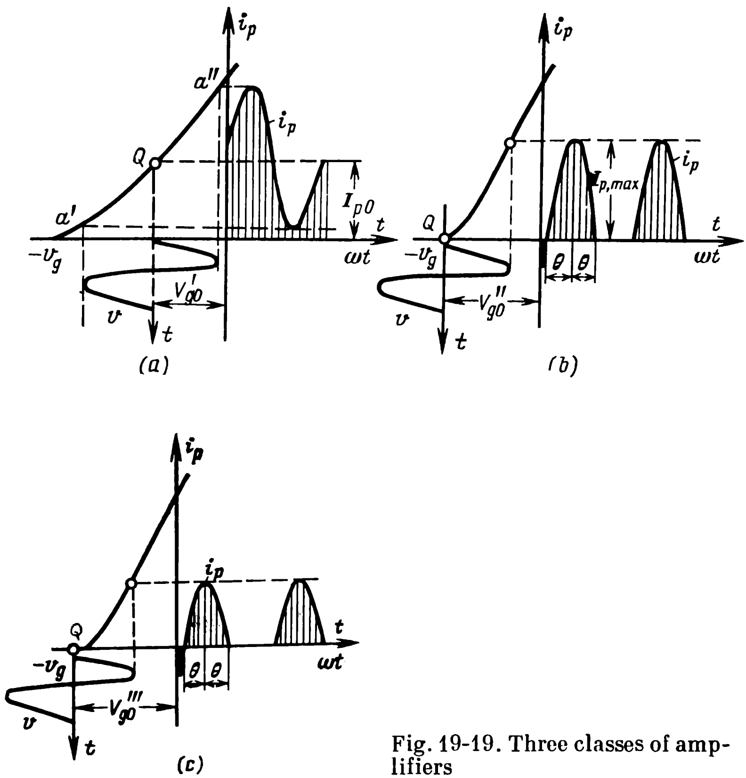


Fig. 19-19. Three classes of amplifiers

In Class B operation, distortion is heavy, but the efficiency is greater than in Class A operation, being 60 to 70%. Class B operation is confined only to the push-pull type of power amplifiers.

If an amplifier operates so that its grid draws no current, the operation is designated Class B_1; if its grid draws current, the operation is designated Class B_2.

Class C (Fig. 19-19*c*). The Q point lies to the left of the grid-plate characteristic, so the grid bias is appreciably beyond the cut-off point, $V''_{g0} > V_{cut\text{-}off}$. In this case, the plate current flows for a fraction of a half-cycle, and the cut-off angle is $\theta < 90°$. In this class, distortion is heavy,

but the efficiency up to 80% and higher. Class C operation is limited to special oscillators and narrow-band amplifiers.

In addition to the three main classes of operation, use is made of intermediate classes.

19-9. Multistage Tube Amplifiers

(a) A Vacuum-Triode Amplifier Stage

Figure 19-20*a* shows the circuit of an amplifier stage built around a vacuum triode, which is often used as a Class A voltage amplifier. This circuit differs from the one considered above (Fig. 19-13*a*) in that it contains an automatic grid bias network, R_kC_k (see Sec. 19-8c). The grid bias V_{g0} is chosen to be somewhat higher than the signal amplitude, $V_{g0} > V_{in,\ m}$, so that $V_{in,\ m}$ would not run beyond the linear region of the grid-plate characteristic or enter the area of positive grid voltage. Owing to this arrangement, distortion is low. In addition to the plate load resistor R_L, the plate voltage is applied to a stage load resistance R_{out}. The blocking capacitor C_b connected in series with R_{out} passes only the a.c. load current and blocks the direct component, because a capacitor offers an infinitely high opposition to direct current. The reactance of the blocking capacitor, $x_{C_b} = 1/2\pi f C_b$, at the operating frequencies of

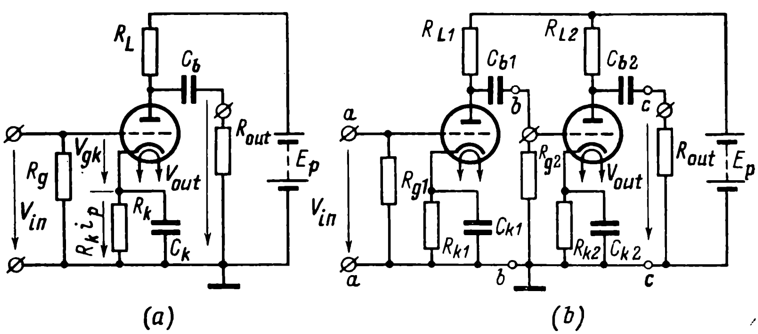


Fig. 19-20. (*a*) Single-stage and (*b*) two-stage amplifier

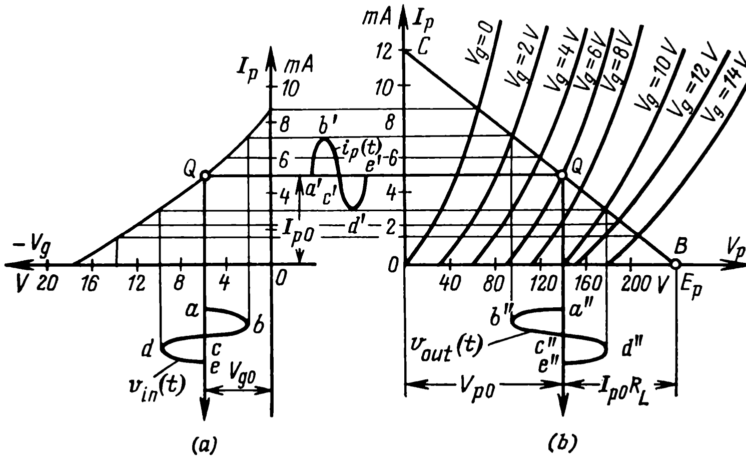


Fig. 19-16. Construction of a stage dynamic transfer characteristic

conductance g'_m of a triode is always less than its static mutual conductance g_m and depends on the plate load resistance. It is defined as follows

$$g'_m = g_m R_p/(R_p + R_L)$$

where R_p is the a.c. (dynamic) plate resistance of the tube.

A dynamic transfer characteristic may be derived from a family of static plate characteristics and the load line (Fig. 19-16). The points where the load line intersects the static characteristics represent the plate current at different values of grid voltage. The grid voltages and the respective plate currents transferred to another coordinate system give the dynamic-transfer characteristic of the stage (Fig. 19-16*a*).

In the no-signal state, the plate current is decided by the position of the Q point on the dynamic-transfer characteristic. In turn, the position of the Q point depends on the grid voltage $V_{g0} = E_g$ which is called the *grid bias voltage* or, simply, *grid bias*. The Q point is positioned in accordance with the required operation class of the amplifier stage (see Sec. 19-8,d). In particular, if the signal is to be amplified without distortion, the bias voltage should not

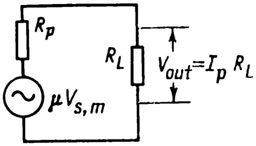


Fig. 19-17. Equivalent circuit for a triode used as a voltage amplifier

be less than the signal amplitude, $E_g > V_{s,m}$; also, the triode should operate within the linear region of its dynamic-transfer characteristic.

Figure 19-16*a* shows the grid bias V_{g0} and the signal voltage $v_s = V_{s,m} \sin \omega t$ (curve *abcde*). The plate current following the signal voltage is given by curve *a'b'c'd'e'* derived from the grid voltage and dynamic-transfer characteristic. The current flowing in the plate circuit develops a voltage drop $v_{out} = i_p R_L$ across the plate load resistor, which is the output voltage of the amplifier stage. In Fig. 19-16*b*, this voltage is represented by curve *a"b"c"d"e"* derived from the plate current waveform and the load line.

The amplitude ratio of the output to input signal voltage gives the stage gain

$$k = V_{R,m}/V_{s,m} \tag{19-19}$$

To correlate the amplification factor of the triode with the stage gain, let us substitute for the amplifier circuit given in Fig. 19-13 its equivalent circuit (Fig. 19-17) where the triode is replaced by an a.c. voltage generator connected in series with the a.c. plate resistance R_p, and peak voltage $\mu V_{s,m}$.

If the generator operates into a plate load resistor R_L, Ohm's law will give the peak current as follows

$$I_{p,m} = \mu V_{s,m}/(R_p + R_L) \tag{19-20}$$

and the peak voltage across the load resistor, which is equal to the peak output voltage, may be written

$$V_{out,m} = I_{p,m} R_L = \mu V_{s,m} \frac{R_L}{R_p + R_L}$$

Hence, the stage gain is

$$k = V_{R,m}/V_{s,m} = \mu R_L/(R_p + R_L) \tag{19-19a}$$

Since $R_L < (R_L + R_p)$, the stage gain is always less than the amplification factor of the tube.

the amplifier should be low as compared with R_{out}, so that all of the a.c. component of the plate voltage is applied to R_{out}.

Then, if a sinusoidal signal voltage, $v_{in} = V_{in,\,m} \sin \omega t$, is applied to the stage input, the amplified voltage at the stage output or, which is the same, across the stage load resistance R_{out}, will be $v_{out} = -kV_{in,\,m} \sin \omega t = -V_{out,\,m} \sin \omega t$. The above expression shows that the output voltage is k times the input voltage in amplitude and the stage output voltage differs a half-cycle or 180° from the input voltage in phase, which is indicated by the "minus" sign.

(b) The RC-Coupled Two-Stage Amplifier

The gain of the single-stage amplifier considered above does not usually exceed a few tens. So whenever a great gain is desired, a multistage amplifier or an amplifier stage built around a vacuum pentode is used.

A two-stage vacuum-triode amplifier (Fig. 19-20*b*) is used to amplify voltage in the frequency range from several hertz to 100 kHz. To reduce distortion, it is usually operated in Class A. This amplifier consists of two identical stages similar to that discussed above.

The stages are coupled by blocking (or coupling) capacitor C_{b1} and grid resistor R_{g2} (Fig. 19-20*b*).

Capacitor C_{b1} prevents the direct current component from flowing via resistor R_{g2} connected at the output of the first stage. Therefore, the voltage at the grid V_{g2} of the second triode does not depend on the direct (supply) current component in the plate circuit of the first tube. However, the blocking capacitor should offer a much lower opposition to the alternating (signal) components (one-tenth or one-twentieth) than R_{g2}. The capacitance of the blocking capacitor usually ranges from thousands to tens of thousands of picofarads. It can be derived from the following relation

$$1/\omega_L C_{b1} \leqslant 0.05 R_{g2} \qquad (19\text{-}22)$$

The input signal voltage, v_{in}, produces a pulsating current in the plate circuit of the first tube. The direct component of this plate current is free to flow through the output

plate load resistor R_{L1}, but it cannot flow through the capacitor C_{b1} to the grid resistor R_{g2}. Some of the alternating component of the first tube plate current flowing through capacitor C_{b1} and grid resistor R_{g2} of the second tube produces across the latter a voltage which is the output voltage of the first stage, $V_{out\,1} = V_{in}k_1$. At the same time, this is the input voltage for the second stage, $V_{out\,1} = V_{in2}$. On being amplified by the second tube, it appears at the output terminals of the second stage as $V_{out\,2} = V_{in2}k_2 = k_1k_2V_{in1}$.

RC-coupled amplifiers have found wide application because they have low signal distortion, good frequency response, and are simple in design, small in size and mass, and low in cost.

Among their main drawbacks are a high voltage drop across the plate load resistor due to the d.c. component of the plate current, and the need for high supply voltage.

19-10. Power Amplifiers

As already noted in Sec. 19-6, an audio-frequency amplifier is usually the final stage of an amplifier unit. Everything that will be said in the first paragraphs of this section applies to both transistor and vacuum-tube amplifiers.

A simplified circuit of an a.f. amplifier stage built around a vacuum triode shown in Fig. 19-21 is self-explanatory. It is stated in Sec. 19-6 that maximum power is delivered to the load when the load resistance matches the internal resistance of the transistor. This requirement is also true of an amplifier stage based on a vacuum triode. So, if the load resistance is such that $R_L = (0.1 \text{ to } 2)\,R_p$, it may be connected directly in the plate circuit.

If a load is such that $R_L \ll R_p$, it should be matched by means of an output transformer (Fig. 19-22) whose turns ratio can be derived from Eq. (19-14)

$$k_t = \sqrt{(0.1 \text{ to } 2)R_p/R_L}$$

Single-ended power amplifiers are used when the required power output does not exceed 3 to 5 W.

A push-pull power amplifier, that is, the one based on two tubes (or transistors) to whose control grids (bases)

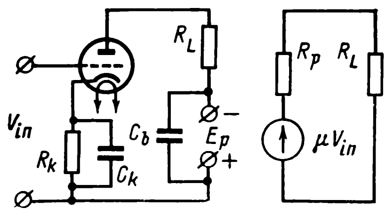


Fig. 19-21. Simplified circuit of an a.f. power amplifier and its equivalent circuit

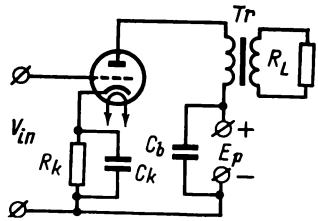


Fig. 19-22. Power amplifier with an output transformer

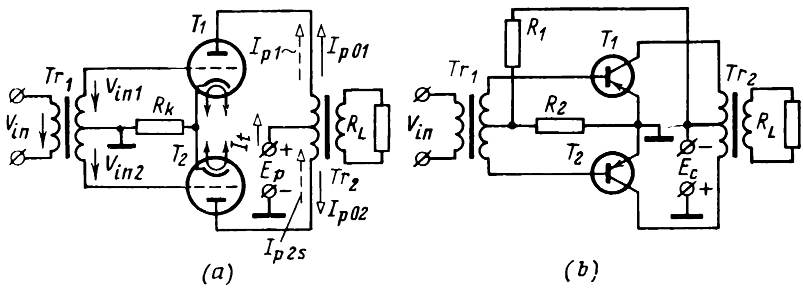
the input voltage is applied in anti-phase, is used when the required power output is over 5 W. The circuits of push-pull output stages using vacuum-tubes and transistors are shown in Fig. 19-23. Both circuits operate in a similar manner.

Each circuit is symmetrical and consists of two equal sections, called arms (in Fig. 19-23*a* and *b*, the upper and the lower arm), corresponding to the circuits shown in Figs. 19-22 and 19-10, respectively. The two similar tubes (transistors) take their plate (collector) power from a common source (E_p or E_c) via two halves of the primary winding of output transformer Tr_2. The bias voltage is applied to the tube grids (transistor bases) through the second winding of input transformer Tr_1. The function of the output transformer is the same as in a single-ended circuit (Fig. 19-22).

Let us consider Class A operation of the tube amplifier (see Sec. 19-5). In this case, the grid bias is taken from the cathode resistor, R_k. In the no-signal state, $V_{in1} = V_{in2} = 0$, the grids are only at bias potential equal to $I_{p0}R_k$, and the quiescent plate currents due to E_p and E_g are equal ($I_{p01} = I_{p02} = I_{p0}$) and flow through two halves of the primary winding of transformer Tr_2 in opposite directions, so that the resultant mmf due to them is zero. Thus, no d.c. bias is applied to transformer Tr_2.

The input voltage (Fig. 19-24*a*) is simultaneously applied to the grids of both tubes in anti-phase:

$$v_{in1} = V_{in,\,m} \sin \omega t, \quad v_{in2} = V_{in,\,m} \sin (\omega t + 180°)$$



ʼFig. 19-23. Push-pull power amplifiers
(a) vacuum-tube ampli er; (b) transistor amplifier

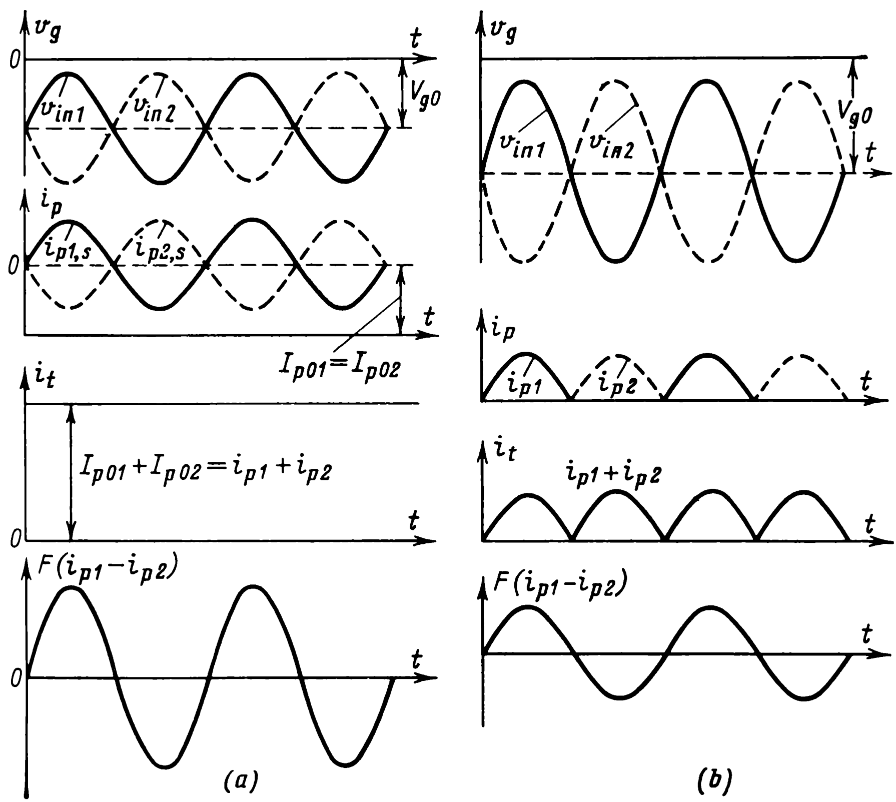


Fig. 19-24. Push-pull amplifier operating Class A
(a) waveforms for grid voltage, plate current and mmf F; (b) waveforms for input voltage, plate current and mmf F' of an amplifier operating Class B

The plate currents which are the sums of direct and alternating components may be written

$$i_{p1} = I_{p01} + i_{p1,\,s} = I_{p01} + I_{p1,\,m} \sin \omega t$$

and

$$i_{p2} = I_{p02} + i_{p2,\,s} = I_{p02} + I_{p2,\,m} \sin (\omega t + 180^\circ)$$
$$= I_{p02} - I_{p2,\,m} = \sin \omega t$$

where $i_{p1,\,s}$, $i_{p2,\,s}$, $I_{p1,\,m}$ and $I_{p2,\,m}$ are the instantaneous and peak values of the alternating plate current.

The total current in the lead connected to the plate source E_p is

$$i_t = i_{p1} + i_{p2} = 2I_{p0}$$

As is seen, it contains no alternating component.

The plate currents i_{p1} and i_{p2} flow through the respective primary half-windings of the output transformer (whose turns ratio is unity) in opposite directions, so, the resultant or load current is

$$i_L = i_{p1} - i_{p2} = 2I_{p,\,m} \sin \omega t$$

Thus, the mmfs due to the alternating plate currents are added together in the output transformer, the magnetic flux is doubled, and so is the power delivered to the load.

Since no d.c. biasing is applied to the output transformer core, its magnetic flux and the output voltage are sinusoidal. In other words, a push-pull amplifier operating in Class A introduces insignificant nonlinear distorion.

A major drawback of Class A push-pull amplifier is low efficiency which is about 15 to 40%.

Push-pull power amplifiers often operate in Class B. As is known from Sec. 19-8 (d), in this case the plate current flows through the triode only for half of each cycle when the grid is positive; during the other half of a cycle, the tube is turned off. In the no-signal state, the plate current is practically zero, so the voltage drop across the cathode resistor due to this current is zero, too. Therefore, a bias voltage source, E_g, is required for this class of operation.

In a push-pull amplifier, the input (signal) voltage is applied to the grids of both tubes. During one half of each cycle, the input voltage is positive at the grid of one tube

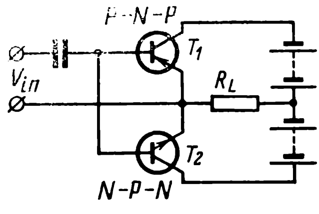


Fig. 19-25. Transformerless push-pull power amplifier built around semiconductor triodes

and negative at the grid of the other. During the next half of each cycle, the grids reverse polarity. As a consequence, the plate current flows through the tubes in turn. If the input (signal) voltage is sinusoidal, the plate currents i_{p1} and i_{p2} are half-sinewaves (Figs. 19-7 and 19-24*b*). As they flow in turn via the two halves of the primary in Tr_2, these currents establish an mmf which excites a practically sinusoidal magnetic flux, and the output voltage is sinusoidal.

In Class B operation, the efficiency of the amplifier is considerably higher than in Class A, being 60 to 70%.

A major drawback of the push-pull amplifier is that it requires two tubes and two centre-tapped transformers.

As already noted, the transistor push-pull power amplifier (Fig. 19-23*b*) operates in much the same manner as that built around vacuum triodes. The transistors are usually connected in a common-emitter circuit. The supply voltages for the emitter and base are taken from a voltage divider, R_1/R_2.

More often transistor push-pull amplifiers operate in Class B where a higher efficiency can be achieved, although nonlinear distortion in this case is heavier than in Class A and increased still more by the asymmetry of the circuit.

In addition to the push-pull amplifiers considered above, use is widely made of transformerless push-pull amplifier stages built around *P-N-P* and *N-P-N* transistors (Fig. 19-25). In the no-signal condition, the transistors are turned off because the base and emitter potentials are equal, which fact ensures Class B operation. When a signal is applied to the input, transistor T_2 conducts only for half of each cycle of this signal, and transistor T_1, only for the other half. Among the advantages of this circuit are simple design and the absence of a transformer. The main drawback is that

the output impedance of the amplifier cannot be properly matched to the load impedance.

19-11. Transistors as Switches

Transistors are widely used as statie switches. Quite appropriately, they are then said ot operate in the switch (or pulse) mobe, When a transistor operates as a switch, it alternates between conduction and nonconduction, that is, the ON and OFF states. For a junction transistor, the collector-emitter resistance, r_{ce}, is several ohms in the ON state and about one megohm in the OFF state.

One of the circuits of a common-emitter transistor switch is shown in Fig. 19-26.

The transistor is turned off when the emitter and collector junctions are biased in the reverse direction. For this purpose, a voltage at which the base potential exceeds that of the emitter ($V_{be} \geqslant 0$) should be applied to the input ol the transistor (Fig. 19-26). The circuit can be driven out of this stable OFF state by applying a negative trigger pulse to the input. If both *P-N* junctions of the transistor are then biased in the forward direction, the transistor will reach saturation; this is the second stable, or ON, state of the circuit.

Figure 19-27 shows the static transfer characteristics of the transistor and its load line, *BC*, which relates the collector current to the collector-to-emitter voltage $I_c =$ $= f\,(V_{ce})$. The collector current rises with increasing input current, I_b, and reaches a maximum, $I_{c,\,max}$, when the base current is $I_{b,\,sat} = I_{b4}$ (with R_c and E_c held constant); this condition corresponds to point *B* in Fig. 19-27. $I_{c,\,max}$ is called the saturation current, $I_{c,\,sat} =$ (approx.) E_c/R_c, because it remains unchanged for any increase in the base current. That is why the transistor is said to be saturated.

As is seen in Fig. 19-27, at $I_{c,\,sat}$, the collector-to-emitter voltage, V_{ce}, is zero very nearly.

It is to be remembered that when a transistor operates as a switch, it dissipates a considerable power only during the transition from one stable state to the other. Therefore, the average power dissipation of the transistor switch is low, and the instantaneous values of the emitter and collector

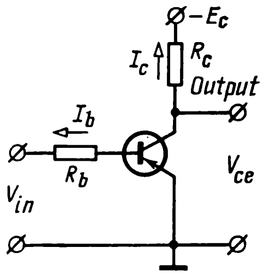


Fig. 19-26. Common-emitter transistor switch

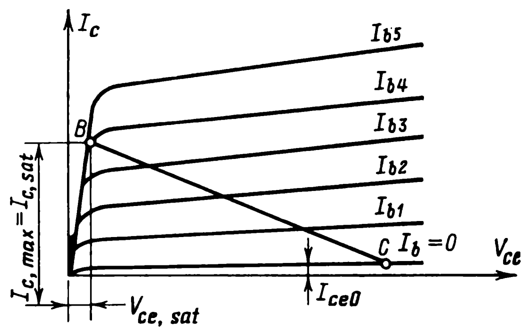


Fig. 19-27. Static transfer characteristics of a transistor and its load line

currents, I_e and I_c, may be several times the values set as limits for amplification.

If the input pulse has a duration much longer than the time of transients (such as charge build-up and decay in the base region), the output pulse will have about the same duration and waveform. If, however, its duration is only a few microseconds, the waveform of the output pulse may be heavily distorted and stretched in time.

Chapter Twenty Electronic Oscillators. Oscilloscopes

20-1. Sinewave or Harmonic Oscillators

In a general sense, an *oscillator* is a device that generates an a.c. power at a desired frequency. It may use one or more tubes or transistors.

High frequency power is predominantly generated by vacuum-tube oscillators.

For its operations, a vacuum-tube oscillator may, or may not, need external excitation. Accordingly, we may have either an *externally excited* or a *self-excited oscillator*. The former is driven by a voltage from an external source, so it is, in effect, a power amplifier. The latter generates oscillations which maintain themselves owing to the resonant properties of the circuit and positive feedback.

In turn, self-excited oscillators may be classed into the *LC* type and the *RC* type, according to the circuit elements used in the resonant circuit and the feedback path*.

(a) LC Oscillators

As has been shown in Sec. 6-8, the frequency of oscillations generated by a vacuum-tube oscillator is determined by the parameters of its *LC* resonant (or oscillatory) circuit, and the tube acts as a regulator which periodically admits energy to the oscillatory circuit, whence it is transferred to load.

The circuit of a likely tube oscillator is shown in Fig. 20-1.

---

* In fact, very few authors class externally excited oscillators as oscillators. Similarly, self-excited oscillators are commonly called simply oscillators or feedback oscillators.-- *Translator's note.*

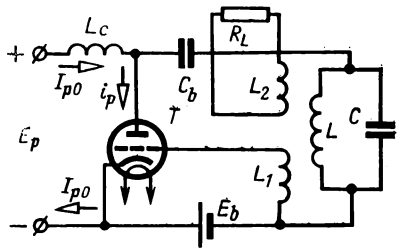


Fig. 20-1. Tube *LC*-oscillator

When the plate source E_p is turned on, the tuned-circuit capacitor C charges to $V_{C,m}$, after which it discharges through the tuned-circuit coil L. This gives rise to oscillations at the natural angular frequency $\omega_0 = 1/\sqrt{LC}$ of the tuned circuit (see Sec. 6-8). There is a feedback coil L_1 connected to the tube grid, so the grid voltage is determined by the emf induced in L_1 at frequency ω_0.

For self-excitation of the oscillator and for the oscillations to sustain themselves two conditions must be satisfied: (1) the voltage applied from the feedback coil to the tube grid must be 180° out of phase with the alternating plate voltage; that is, feedback must be positive; (2) the feedback must be sufficient for the alternating plate current to compensate for losses in the tuned circuit. When these conditions are met, a pulsating current, i_p, decided by the bias voltage E_b, arises in the plate circuit (Fig. 19-9*a*). The direct plate current, I_{p0}, is prevented from entering the tuned circuit by blocking capacitor C_b; instead its path is completed through the power source and r.f. choke L_c. The alternating (or r.f.) plate current, $I_{p,m} \sin \omega_0 t$, cannot reach the power source, because L_c offers a high reactive impedance, $\omega_0 L_c$; so its path is completed through the tuned circuit. Since this current is in phase with the tuned-circuit voltage, power is periodically transferred to the tuned circuit.

The load circuit consists of a load resistor, R_L (Fig. 20-1) and a coil L_2 inductively coupled with the tuned-circuit coil L. Thus, the tuned circuit transfers power to the load via the magnetic flux which links L_2 and L.

In the circuit of Fig. 20-1, feedback between the tuned and grid circuits is accomplished by mutual inductance. Instead of mutual inductance, feedback can be accomplished

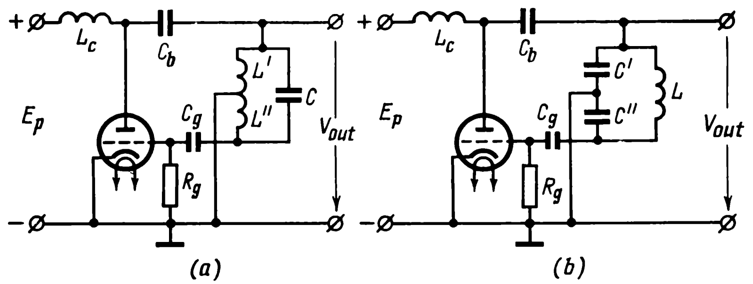


Fig. 20-2. The circuit of (*a*) a Hartley oscillator and (*b*) a Colpitts oscillator

by a tapped-coil or a tapped-capacitor arrangement. The respective oscillators are known as the *Colpitts* and the *Hartley oscillators.* Figure 20-2*a* shows the circuit of a Hartley oscillator, and Fig. 20-2*b* that of a Colpitts oscillator.

The two conditions of oscillators defined above hold for these oscillators as well. The voltage fed back to the grid at the proper amplitude and in the proper phase is taken from coil L'' in the Hartley circuit, and from capacitor C'' in the Colpitts circuit.

All other circuit components serve the same purpose as they do in Fig. 20-1.

(b) RC Oscillators

Vacuum-tube *LC* oscillators are mainly used at frequencies outside the audio range, that is, above 20 kHz. In the audio range, use is more often made of simpler, less expensive and more convenient resistance-capacitance (*RC*) oscillators (Fig. 20-3).

Instead of a tuned circuit, an *RC* oscillator uses a load resistor, R_L, and feedback is accomplished by a multistage *RC* network.

Consider a simplified voltage vector diagram for an *RC* network made up of three resistors and capacitors (Fig. 20-4). For simplicity, the current in each next *RC* section may be neglected in comparison with the current in the previous section. Then the plate voltage V_p applied to the first sec-

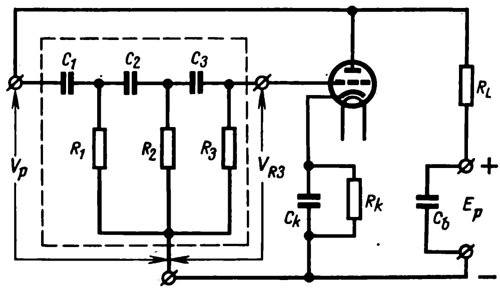


Fig. 20-3. *RC* oscillator

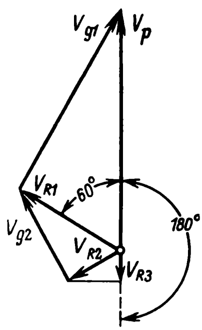


Fig. 20-4. Voltage vector diagram for three-section *RC* network

tion, R_1C_1, will consist of two components, namely the capacitor voltage, V_{C1}, leading the current by 90°, and the resistor voltage, V_{R1}, in phase with the current and leading V_p by an angle φ_1. By adjusting the values of R_1 and x_{C1} and the frequency, this angle may be made equal to 60°.

The resistor R_1 is connected to the second section, R_2C_2. The resistor voltage, V_{R2}, likewise leads V_{R1} by $\varphi_2 = 60°$. Similarly, V_{R3} leads V_{R2} by $\varphi_3 = 60°$. Thus, at a certain frequency, the three-stage *RC* network discussed above produces a 180° phase shift between the output and input voltages, V_{R3} and V_p.

So, in the circuit of Fig. 20-3, the grid voltage $V_g = V_{R3}$ is in antiphase with the plate voltage, V_p, thereby satisfying the condition for oscillation. The frequency of oscillations is defined by the parameters of the feedback network

$$f_0 = 1/2\pi\sqrt{6}RC = 1/15.4RC \qquad (20\text{-}1)$$

The frequency can be changed by adjusting the values of R's and C's in the network.

It is an easy matter to prove that the voltage across R_3, when applied to the grid, is 1/29 of the voltage across the first section, R_1C_1, that is, $V_g/V_p = 1/29$. When the amplifier gain is $k = 29$, oscillations are sinusoidal; when $k > 29$, they are nonsinusoidal; when $k < 29$, no oscillations are generated.

20-2. The Sawtooth Voltage Generator

A sawtooth voltage generator is a relaxation oscillator. Relaxation oscillators are devices which generate non-sinusoidal repetitive waveforms.

A sawtooth voltage is a voltage which rises to its peak value at a relatively low rate with time and falls quickly (almost instantaneously) back to its initial value, so that its waveform resembles the teeth of a saw. Figure 20-5 represents an ideal and a real sawtooth waveform.

A sawtooth voltage can be obtained by causing a capacitor to charge slowly and discharge rapidly.

The basic sawtooth generator (Fig. 20-6) consists of a capacitor C, neon tube connected across the capacitor, and a resistor R_a connected in series with the shunt arm.

If we apply the direct anode-supply voltage, V_a, to the oscillator, the capacitor will charge through resistor R_a, and its voltage will rise exponentially [see Eq. (18-22)] during the first part of a cycle, T_r (Fig. 20-5)

$$v_C = V_a [1 - \exp(-t/\tau_1)]$$

until it reaches V_f, the firing voltage of the neon tube. At that instant, the tube resistance suddenly drops and the capacitor discharges quickly through the tube, so that its voltage decreases exponentially [see Eq. (18-24)]:

$$v_C = V_{C,\,m} \exp(-t/\tau_2)$$

When the capacitor voltage falls to the extinction voltage, V_e, of the tube, the tube ceases to conduct, the capacitor begins to re-charge, and the cycle just described starts all over again.

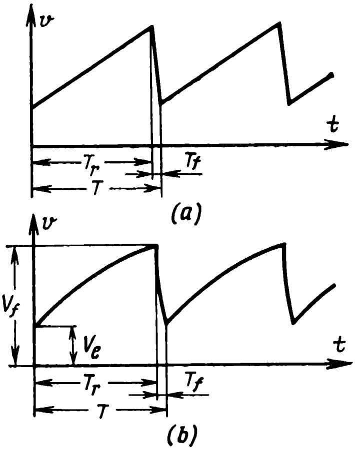


Fig. 20-5. (*a*) Ideal and (*b*) real sawtooth voltage waveforms

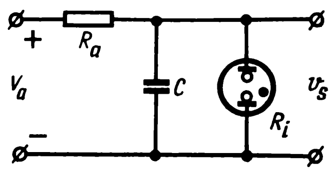


Fig. 20-6. Sawtooth voltage generator using a neon tube

The rate at which the capacitor charges and discharges is characterized in terms of the time constant $\tau = RC$. By adjusting the time constant, we can control the period (or duration) of the sawtooth voltage cycle. This can be done by adjusting either the resistance or the capacitance, or both.

This type of oscillator is used relatively seldom, because it has a number of disadvantages. In particular, the firing and extinction voltages of the neon tube are unstable, and the difference between the firing and extinction voltages $\Delta V = V_f - V_e$ is insufficient.

A better sawtooth generator can be built, if we replace the neon tube by a thyratron or a vacuum-tube circuit.

The circuit of a sawtooth generator based on a thyratron is given in Fig. 20-7.

A typical thyratron has a low extinction voltage (about 15 V) and its firing voltage can be controlled by grid bias, so the amplitude of the resultant sawtooth voltage can be raised appreciably. A particular duration range for the sawtooth voltage is selected with switched capacitors C_1, C_2, C_3, etc., and within each range it can be adjusted continuously by variable resistor R_a. The operation of the thyratron can

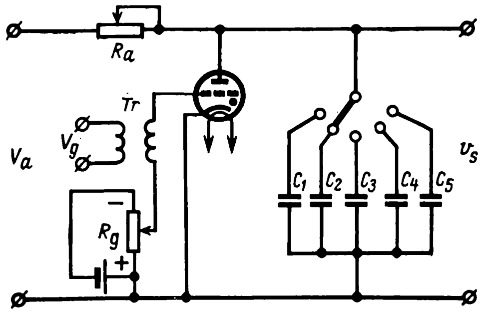


Fig. 20-7. Sawtooth voltage generator based on a thyratron

conveniently be synchronized with any voltage source, owing to a transformer in the grid circuit.

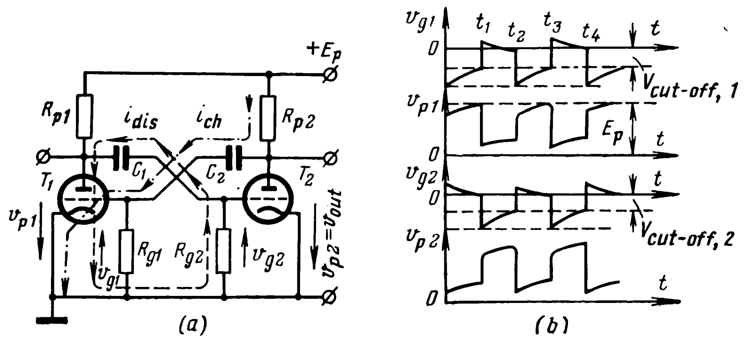
20-3. Multivibrators

A multivibrator is a relaxation oscillator generating square or rectangular voltage pulses by using the charging and discharging of capacitors. A multivibrator may be regarded as a two-stage amplifier (Fig. 20-8*a*) with positive feedback such that the output (plate) voltage of the first stage is applied to the input (grid) of the second stage, and the output voltage of the second stage is applied to the input of the first. Owing to this feedback, the tube of the first stage and that of the second stage are instantaneously rendered conducting and nonconducting in turn and in a rapid succession.

If a multivibrator uses identical tubes, T_1 and T_2, identical grid resistors, R_{g1} and R_{g2}, identical plate resistors, R_{p1} and R_{p2}, and identical capacitors, C_1 and C_2, it is called a *symmetrical multivibrator* (Fig. 20-8*a*).

Since both stages of the multivibrator are identical, the tubes, T_1 and T_2, will draw identical currents immediately after power is turned on, as one should expect. However, this state of the circuit is unstable.

Assume that the plate current i_{p1} of the first tube has increased for some reason. This will give rise to a voltage drop across R_{p1}, and, as a consequence, the plate voltage,



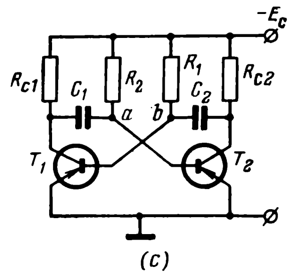


Fig. 20-8. (*a*) Symmetrical vacuum-tube multivibrator; (*b*) its voltage waveforms; (*c*) transistor multivibrator

v_{p1}, of this tube will go down. This decrease in the plate voltage will entail a decrease in the voltage across the series $R_{g2}C_1$ network. However, the voltage across a capacitor cannot change instantaneously, so the voltage decrease is transferred to R_{g2}, causing the grid potential of T_2 to go down, its plate current i_{p2} to fall, and its plate voltage v_{p2} to rise. This voltage rise is transferred to the $R_{g1}C_2$ network, so the voltage across resistor R_{g1} and at the grid of T_1 builds up. The rise in v_{g1} leads to a still greater increase in the plate currents of T_1. The process is complete when T_1 is conducting heavily and T_2 is driven beyond cut-off—the multivibrator is said to have changed its state; this change of state proceeds cumulatively and occurs in a fraction of a microsecond. In Fig. 20-8*b*, the multivibrator changes state at time t_1.

As soon as T_2 is driven beyond cut-off, the capacitor C_1 charged to the plate supply voltage, E_p, will discharge with

a high time constant, τ_{dis}, through T_1 (its a.c. plate resistance, R_{p1}), and resistor R_{g2}. As the capacitor discharges, the voltage drop across R_{g2} due to the discharge current goes down, so the negative voltage at the grid of T_2 decreases. When this voltage rises just above cut-off, T_2 will be rendered conducting, and a current, i_{p2}, will begin flowing.

Just as C_1 begins discharging, that is, just as T_2 is turned off and its plate voltage begins rising, C_2 will start to recharge. The charge current i_{ch} will flow from $+E_p$ through R_{p2}, C_2 and the shunt arm consisting of R_{g1} and the grid-to-cathode space of tube T_1, to $-E_p$. The charge current develops a voltage drop across R_{g1}, so that the grid voltage of T_1 increases, the current i_{p1} rises still further, and the plate voltage v_{p1} goes down. Since the discharge time constant is much longer than the charge time constant ($\tau_{dis} > \tau_{ch}$), C_2 stops charging long before T_2 is driven to conduction.

Just as T_2 is rendered conducting, the plate voltage v_{p2} decreases, and this decrease is transferred via the $R_{C1}C_2$ network to the grid of T_1; as a result, i_{p1} goes down and i_{p2} builds up. This process likewise proceeds cumulatively and the circuit jumps to a new state where T_1 is driven beyond cut-off and T_2 is conducting heavily. In Fig. 20-8*b*, this occurs at t_2.

This state is unstable, too, and the circuit will keep changing states until the plate source E_p is open-circuited.

Just as the tubes are periodically turned on and off at times t_1, t_2, t_3, . . ., nearly rectangular voltage pulses appear at the output of each tube (Fig. 20-8*b*).

The frequency of the multivibrator can be adjusted by varying the resistances of R_{g1} and R_{g2} or the capacitances of C_1 and C_2.

In addition to vacuum-tube multivibrators, use is made of transistor multivibrators. The circuit of a free running transistor multivibrator is in Fig. 20-8*c*.

This operates on much the same principle as the circuit based on vacuum tubes. It consists of two identical stages (a symmetrical multivibrator). The output of the first stage is connected to the input of the other and vice versa. In contrast to the circuit considered above, the multivibrator of Fig. 20-8*c* has its feedback resistors R_1 and R_2 returned

to the "—" side of the supply ($-E_c$) rather than to chassis ground.

Transistors T_1 and T_2 are rendered conducting in turn because capacitors C_1 and C_2 charge also in turn. For example, when transistor T_1 is conducting for a given part of a cycle, the source $-E_c$ charges capacitor C_1; and the charge current flowing through R_{c1} and the base-emitter junction of the conducting transistor T_2 develops a voltage, v_{C1}, across C_1. At the same time, C_2 discharges via R_1 and R_{c2}, and the voltage drop across R_1 caused by this discharge current produces a positive potential at point b or, which is the same, at the base of T_1, so that T_1 is turned off.

As C_2 discharges, its discharge current decreases exponentially, and so do the voltage drop due to this current across the resistor and potential at point b. When the potential at point b falls nearly to zero, T_1 turns on suddenly (in a fraction of a microsecond) due to the effect of the feedback circuits, and T_2 turns off. After T_1 has been driven to conduction, C_1 begins to discharge and C_2 to charge. Then the chain of events is all over again.

As compared with vacuum-tube multivibrators, transistor multivibrators have higher efficiency and require smaller supply sources, but their parameters are markedly affected by temperature variations.

20-4. Cathode-Ray Tubes

The cathode-ray tube is a vacuum tube in which its electron beam can be focused to a small cross section on a luminescent screen and can be varied in position and intensity to produce a visible pattern from electric signals.

By their purpose, cathode-ray tubes may be divided into three basic groups:

—oscilloscopic cathode-ray tubes. They are used in the study of fast-varying repetitive and nonrepetitive processes;

—display or indicator cathode-ray tubes. As their name implies, they are used to display signals or serve in indicators;

—TV receiver cathode-ray tubes (picture tubes or kinescopes).

A cathode-ray tube (or CRT, for short) consists essentially of:

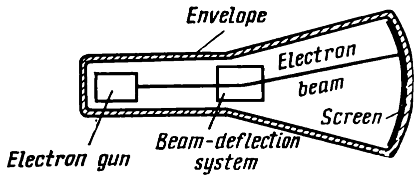


Fig. 20-9. Construction of a CRT

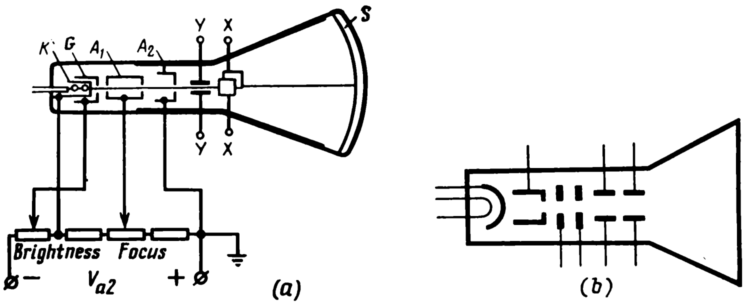


Fig. 20-10. Electrostatic CRT
(*a*) electrode structure and supply; (*b*) diagram symbol

Fig. 20-11. Cathode of a CRT

(1) an evacuated glass envelope (Fig. 20-9);

(2) an electrode structure, called the electron gun, which produces an electron stream and forms an electron beam;

(3) a beam-deflection system intended to position the electron beam on the screen;

(4) a screen which glows when struck by electrons.

The beam can be positioned electrostatically or magnetically.

Consider the design of an electrostatic CRT (Fig. 20-10).

The envelope is rather long, funnel-shaped bulb in which a high vacuum is produced. In its neck is located the electron gun—a combination of an indirectly-heated oxide-coated cathode, *K* (sometimes called the emitter), a control electrode or grid, *G*, and anodes, A_1 and A_2, which focus the beam.

The cathode (Fig. 20-11) has a tungsten heater, *1*, located inside a nickel cylinder, *2*, whose flattened end, *3*, is given an outside coat of oxide, *4*, so that electrons are emitted in one direction and from one end only.

The cathode is enclosed in the control grid, *G* (Fig. 20-10), which is a cylinder with a small aperture in the bottom. It controls, or modulates, the electron beam emitted by the cathode. The control grid is held at a small negative potential relative to the cathode.

Under the action of the electric field between the cathode and control grid, the electrons leaving the cathode at, say, point *a* in the direction *aa′* (Fig. 20-12*a*) change their path and move in the direction *bc*, that is, towards the axis of the beam. If the control grid is made more negative, some of the electrons will be deflected more and miss the aperture. Thus, by varying the control grid potential, we can control the number of electrons in the beam and the brightness of the spot on the screen.

On leaving the control grid, the electrons tend to spread or move away from the beam axis. This tendency is counteracted by the anodes A_1 and A_2 (Fig. 20-12*b*). The first (or focus) anode is a cylinder enclosing two or three aperture discs; the second (or accelerating) anode is also a cylinder enclosing one or no aperture disc. The first anode is held at a positive potential of 0.2 to 0.5 kV, and the second at a positive potential of 1 to 2 kV relative to the cathode. The electrons entering the electric field between these anodes are deflected towards the beam axis (focused) and accelerated. This is illustrated in Fig. 20-12*c* which shows an electric line of force in the field between the anodes A_1 and A_2 and the electron trajectory, *ABB′A′*. In points *B* and *B′*, where this trajectory intersects the line of force, two forces,

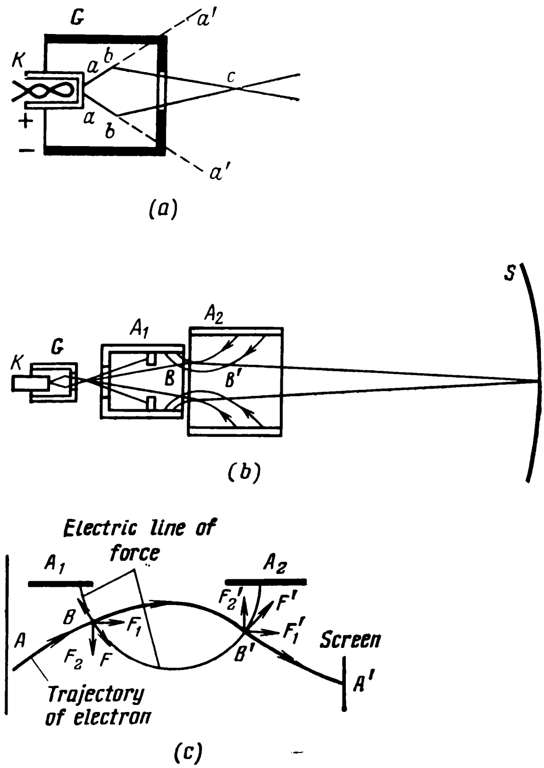


Fig. 20-12. (*a*) Cathode and control grid; (*b*) focusing of an electron beam; (*c*) trajectory of an electron moving in the field between anodes A_1 and A_2

F and F_1, are shown. These forces act on the moving electron in directions tangent to the line of force at those points. Each force can be resolved into two components, longitudinal (F_1 and F_1') and transverse or radial (F_2 and F_2'). The longitudinal components accelerate the electrons along the tube axis, and the transverse components deflect the electrons towards or away from the tube axis (F_2 and F_2', respectively). Thus, the field between the anodes acts like a converging lens focusing the electron beam at a point on the tube axis near or on the screen, S (Fig. 20-12*b*). As a result, a very

small light spot will be seen on the screen. The size of the spot can be controlled by varying the potentials at the first anode.

In more advanced electron guns used in present-day oscilloscopic CRTs, the control grid is separated from the first anode by an accelerating electrode—a metal disc with an aperture in the centre—which is held at a very high positive potential. Owing to this arrangement, the brightness and focus controls no longer interfere with each other, and the operating conditions of the tube can be controlled more effectively.

On striking the screen, the electrons give up their energy which is partly converted into light and partly imparted to the electrons of the screen phosphor, thereby initiating secondary emission. The secondary electrons are collected by a conductive coating applied to the inside of the tube's neck and funnel-shaped portion and connected to the accelerating anode. Most often, the material of this conductive coating is aquadag, a lubricating compound containing graphite.

As already noted, the deflection system is intended to position the electron beam on the screen. It consists of two pairs of deflection plates located at right angles to each other (Fig. 20-10).

One pair, Y-Y, deflects the beam vertically, so they are called the vertical-deflection (or Y-) plates. The other pair, X-X, deflects the beam horizontally, and they are called the horizontal-deflection (or X-) plates.

The electric field established between each pair causes the electron beam to deflect in the respective direction. Assume that the beam coincides with the axis of the tube, in which case the light spot will appear in the centre of the screen. When a potential difference, V, is produced across any pair of plates (Fig. 20-13), the electric field established between them causes the beam to deflect, and the spot will take up a different position on the screen. The distance between this position and the tube axis is

$$h = 0.5\,(V/V_a)\,(l/b)\,(L + l/2) = \text{(approx.)}\ 0.5\,(V/V_a)\,(l/b)\,L$$

where V_a is the anode potential.

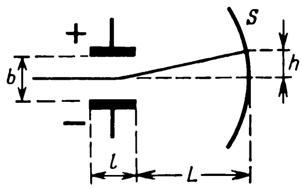


Fig. 20-13. Deflection of an electron beam by an electric field

The displacement of the electron beam on the screen per volt of deflection potential is called the *deflection sensitivity of a CRT*:

$$S_V = h/V = 0.5\ Ll/V_a b$$

For most CRTs, it ranges between 0.2 and 0.6 mm/V.

The ability of some materials to emit light when struck by electrons is called *cathodoluminescence*, and the materials are called *phosphors*. The phosphors of CRTs intended for visual observation are made from either artificially prepared zinc silicate (Zn_2SiO_4) or the naturally occurring mineral willemite, activated with manganese. They give off yellow-green fluorescence. CRTs for oscillography use calcium tungstate ($CaWO_4$) which produces a glow of blue colour.

An important parameter of a phosphor is its *persistence* (*or decay*) *characteristic* defined as the time interval between the instant when the electron beam is turned off and the instant when the intensity of fluorescence has fallen to one per cent of its initial brightness.

Long-persistence CRTs use two-layer phosphors. The layer next to the glass is a long-persistence photoluminescent material (for example, zinc sulphite), and the topping layer is a cathodoluminescent material.

The CRT screen is round and occupies all of the front-end surface of the funnel-shaped part of the tube. The light pattern on the screen is viewed on from the side opposite to that excited by electrons. Therefore, the thickness of the phosphor should exceed the penetration depth of the beam, but, at the same time, it should be small enough, so that the layer next to the glass absorbs light as little as possible.

20-5. The Cathode-Ray Oscilloscope

A CRT oscilloscope (or a CRT oscillograph) is an electronic instrument intended to display (or to record) an electrical quantity (voltage, current, etc.) as a function of time.

As its name implies, the main part of an oscilloscope is an electrostatic cathode-ray tube.

The voltage, v, whose waveform is to be viewed (or recorded), is applied to the Y-plates of the CRT, and to the X-plates is applied a sweep voltage, v_s, usually having a sawtooth waveform, adjusted to have a period equal to, or a multiple of, that of the unknown voltage, or signal (Fig. 20-14*a*, and *b*).

Initially (at t_1), the sweep voltage is zero. Gradually, it rises with time so that it reaches its peak value, $V_{s,\,m}$, at the end of a cycle of the signal. In doing so, it forces the electron beam to move across the screen from point a_1 to point a_5 at a constant speed (Fig. 20-14).

If, at the same time, it is acted upon by the electric field established between the Y-plates by the signal voltage, the beam will be deflected vertically for a distance proportional to the instantaneous value of the signal. Owing to the joint action of the sweep and signal voltages, the beam traces out a waveform which relates variations in the signal to time (Fig. 20-14*c*).

On reaching its peak at t_5, the sweep voltage instantaneously falls down to zero and, as a consequence, the electron beam flies back from point a_5 to point a_1 in a straight line. Then, the sweep voltage again rises with time, and the electron beam re-traces its path during the second and all subsequent cycles.

The persistence of the phosphor enables the CRT to display a stationary pattern representing the unknown quantity, provided that the period of the sawtooth voltage is adjusted to a multiple of that of the signal. For example, if the ratio of the two voltages is n, the CRT will display n cycles of the signal waveform.

In practice, the sweep voltage differs from the ideal sawtooth waveform considered above. In particular, the re-trace (or fly-back) part of the sawtooth wave is sloping rather than vertical, because the sweep voltage falls from its peak

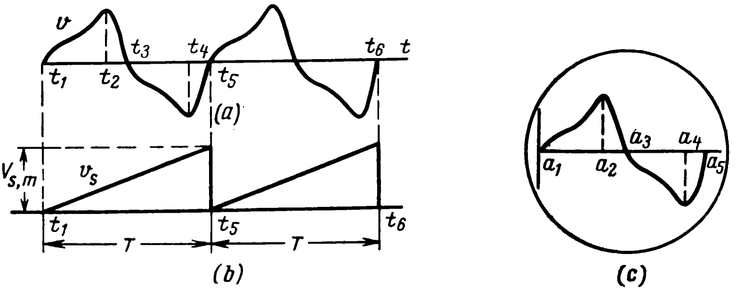


Fig. 20-14. Waveforms for the unknown voltage v and sweep voltage v_s

to zero during a small part of a cycle rather than instantaneously. Accordingly, the respective part of the signal waveform is missed. The retrace of the beam from point a_5 to point a_1 is faster than the forward motion (or scan). Besides, the beam is blanked during retrace, so that it cannot be seen on the screen. The sawtooth voltage is supplied by a sawtooth voltage generator such as discussed in Sec. 20-2.

The deflection sensitivity of both plate pairs is low, so the signal and the sawtooth voltage are amplified before they are applied to the deflection plates. CRT oscilloscopes have two amplifier channels, one for the Y-plates and the other for the X-plates. As a rule, these are multistage vacuum-tube amplifiers.

A block-diagram of an electrostatic CRT oscilloscope is shown in Fig. 20-15.

The electrodes of the electron gun are energized by a rectifier with $E_1 = 1$ to 2 kV via a voltage divider, r_1/r_2. Power for the deflection plates is taken from a rectifier supplying $E_2 = 0.2$ to 0.3 kV. The light spot is initially set on the screen by adjusting the potentials at the deflection plates.

The signal and the sweep voltage are applied to the deflection plates via coupling capacitors. If two repetitive events are to be watched at a time, use is made of a dual-beam oscilloscope or a dual-trace oscilloscope. In a dual-beamoscilloscope, the two signals are displayed at the same time. In a dual-trace oscilloscope, an electronic commutator applies two signals to a single channel in turn;

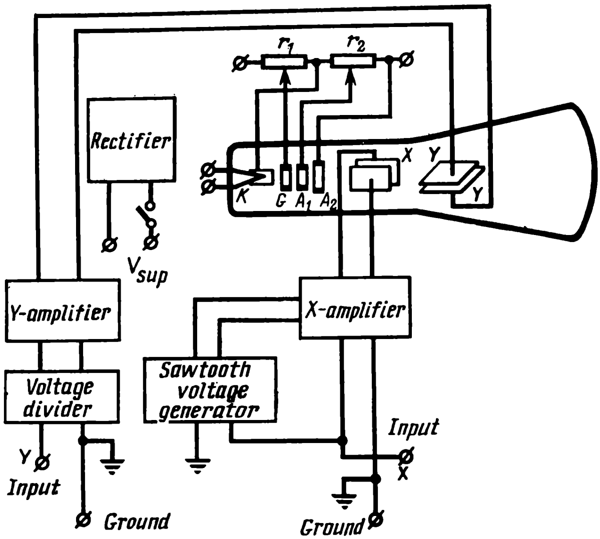


Fig. 20-15. Block diagram of a CRT oscilloscope

yet, with a long-persistence tube both waveforms can be arranged to be seen simultaneously.

CRT oscilloscopes are used not only to display voltage waveforms, but also to measure voltage, current, frequency, power factor, time intervals, and other electrical quantities. Besides, they are used as null indicators; in conjunction with transducers, they can be used to measure nonelectrical quantities.

If we apply the waveform to be viewed to only one pair of deflection plates of an oscilloscope, the screen will present a straight line whose length is proportional to the peak-to-peak amplitude of the signal. By measuring the length of this line, we can determine the voltage amplitude, provided that the voltage constant of the oscilloscope in this mode of operation is known.

Current is measured by applying the voltage developed by the unknown current across a standard resistor to one pair of deflection plates. Having determined the voltage amplitude as already explained, and knowing the value

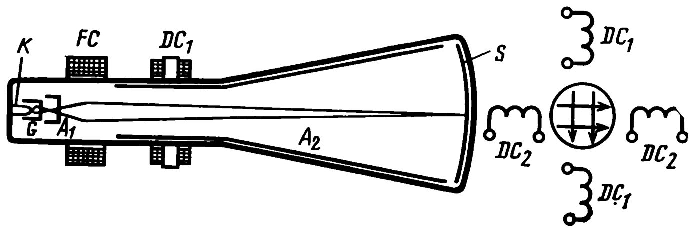


Fig. 20-16. Magnetically focused CRT

of the standard resistor, we can easily find the current amplitude by Ohm's law.

A magnetically focused, magnetically deflected CRT (Fig. 20-16) has a cathode, K, a control grid, G, and an anode, A_1, that have the same design and serve the same purpose as those in an electrostatic CRT.

The electrons are accelerated along the axis by the field due to a second (accelerating) anode, A_2, which is a layer of graphite on the inside of the neck and funnel-shaped part of the tube. The beam is focused by the magnetic field due to the direct current flowing in a focusing coil (or focuser), FC. Because the field inside the coil is nonuniform, the electrons are deflected towards the tube axis and converged (or focused) to some point on the screen as a light spot. The spot size can be controlled by adjusting the current in the focuser.

Magnetic deflection of the beam is effected by a deflection yoke made up of two pairs of deflecting coils, each pair placed at right angles to the other. The deflecting coils, DC_1 and DC_2, may or may not have ferromagnetic cores. The sweep voltage is supplied by a sawtooth generator.

20-6. Coding System of Soviet-Made CRTs

The first element of the CRT designation is the number indicating the screen diameter (or diagonal) in centimetres.

The second element consists of two letters indicating the design of the tube (Russian letters throughout):

ЛО stand for an electrostatic CRT;

ЛМ stand for a magnetic CRT.

The third element is the number which indicates the type of the tube.

The fourth element is the letter indicating the type of phosphor:

P, blue-violet fluorescence and phosphorescence, medium persistence;

И, yellow-green fluorescence and phosphorescence, medium persistence;

У, light-green fluorescence and phosphorescence, short persistence;

A, blue fluorescence and phosphorescence, short persistence;

M, blue fluorescence and phosphorescence, short persistence;

B, blue fluorescence, yellow phosphorescence, long persistence;

K, pink fluorescence, orange phosphorescence, long persistence;

C, orange fluorescence and phosphorescence, long persistence;

Д, blue fluorescence, green phosphorescence, long persistence.

For example, the designation "13ЛО5А" reads as follows: "electrostatic CRT, type five, 13-cm screen diameter, blue colour, short persistence".

Chapter Twenty One

Electron-Tube, Solid-State and Photoelectric Relays

21-1. General

A relay is an automatic device which operates when an input or actuating quantity, x, reaches a predetermined value, so that an output or controlled quantity is caused to change in magnitude abruptly. Referring to the characteristics of a relay (Fig. 21-1), so long as the input quantity, x, changes from zero to x_2, the output quantity, $y = y_1$, remains unchanged. Just as the input quantity reaches the value $x = x_2$, known as the *operate* (or *operating*) *value*, the relay causes the output quantity to change in value from y_1 to y_2. Any further change in x will not affect y.

So long as the input quantity decreases to x_1, the output quantity remains the same, y_2. Just as the input quantity reaches the value x_1, known as the *reset* (or *resetting*) *value*, the relay causes the output quantity to go down abruptly to the value y_1 which is maintained until x becomes zero.

In accordance with the input or actuating quantity used, relays can be classed into current relays, voltage relays, timing relays, etc.

Figure 21-1 represents the characteristic of a relay in which the input quantity is current, $x = I$, and the output quantity is voltage, $y = V$.

The ratio between the reset (resetting) and operate (operating) values of the actuating quantity is called the *reset (resetting) ratio of a relay*. This factor

$$k_r = x_1/x_2 = I_r/I_o$$

ranges from 0.3 to 0.95.

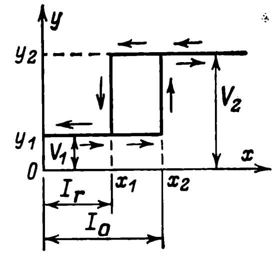


Fig. 21-1. Characteristic of a relay

Another classification of relays is into *movable-contact* and *static types*. Movable-contact relays are less reliable because the contacts tend to fail and wear out rapidly. Static relays are more accurate, more sensitive, faster in operation, more reliable, and last longer in service because they have no moving parts. Owing to this, static relays are finding an ever increasing field of application.

Movable-contact relays are mainly electromechanical relays (Sec. 11-13) which are designed to interpret input conditions in a prescribed manner and, after specified conditions are met, to respond to contact operation in associated electric output (controlled) circuits. This group also includes electronic movable-contact relays which are combinations of electromechanical relays and vacuum-tube or transistor amplifiers.

Examples of static relays are vacuum-tube or transistor switching devices. Very often, they are built as trigger or flip-flop circuits.

Relays are characterized by: (1) the *operate time* t_0, defined as the time interval from the instant when an input quantity is applied to a relay to the instant when the relay operates upon the controlled circuit. For electronic static relays, this time is about 10^{-8} s; (2) the *actuating power*, P_{act}, defined as the electrical power that must be applied to input circuit for a relay to operate positively. For electronic relays, it ranges from 10^{-6} to 10^{-10} W; (3) *sensitivity* or *just-operate value*, defined as the minimum value of the input quantity (usually current) at which the relay operates; (4) *reliability* or *failure-free operation.*

21-2. The D.C. Vacuum-Tube Movable-Contact Relay

Figure 21-2 shows the circuit of a vacuum-tube movable-contact relay. The first (left-hand) triode is driven by a voltage applied between its grid and cathode; it is the sum of three components: input voltage V_{in}, negative grid bias V_b taken from the right-hand section of resistor R_1, and controlled feed-back voltage taken from resistor R_2.

When the input voltage is positive, $V_{in} > 0$ (a positive pulse), the negative voltage, V_{g1}, at the grid of the first triode goes down, the plate current increases, and the plate voltage, V_{p1}, decreases. The voltage applied between the grid and cathode of the second (right-hand) triode is

$$V_{g2} = V_{p1} - I_2 R_2$$

If the plate voltage, V_{p1}, is low, the plate current, I_{p2}, will be smaller than the operate current of relay *Rel* and, consequently the relay contacts will be open.

A decrease in the input voltage V_{in} leads to a rise in the negative grid bias of the first triode and, therefore, to a decrease in its plate current. This causes the plate voltage V_{p1}, and grid voltage, V_{g2}, to increase. As a result, the second triode is rendered conducting, its plate current I_{p2} reaches the operate current value, and the relay contacts

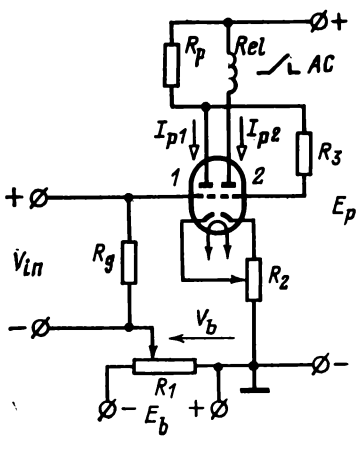


Fig. 21-2. D.c. vacuum-tube movable-contact relay

make. At a certain value of V_{in}, the first triode ceases conducting, and a stable state sets in with the second triode turned on. When the input voltage reaches the value V''_{in}, the first triode is driven to conduction and the circuit is reset. Thus, when the input voltage rises or goes down, the circuit is transferred from one stable state to the other. In one of these states, the actuator circuit (*AC*, Fig. 21-2) is closed, in the other one, it is opened.

21-3. The A.C. Vacuum-Tube Movable-Contact Relay

When an alternating voltage, V_{in}, is applied to terminals *1*, *1'* of the circuit shown in Fig. 21-3, the crystal diode (rectifier), D, and the capacitor, C, produce a direct voltage across terminals *2*, *2'*.

At a low filament voltage, the anode current of a tungsten-cathode vacuum diode, T_1, depends little on the anode voltage V_p, but it rises greatly with increasing filament voltage.

The double triode, T_2, operates as a d.c. amplifier. R_3 is the load resistor of the left-hand triode. The grid voltage of the right-hand triode, which controls the release voltage of the relay, *Rel*, can be adjusted by a potentiometer, R_4, and the operate voltage by a rheostat, R_1.

An increase in the input voltage brings about a rise in the diode filament voltage and, as a consequence, an increase in the current flowing through diode T_1. On the other hand, the anode current, I_p, does not depend on variations in the anode voltage, V_p (the diode is saturated). So, the voltage drop, I_pR_2 across resistor R_2 builds up, and the anode voltage of T_1 goes down. This brings about an increase in the negative grid bias of the left-hand triode which is equal to the difference between the voltage drop across resistor R_3 and the anode voltage of diode T_1. With time, the plate current of the left-hand triode and voltage drop across resistor R_3 decrease. The negative grid bias of the right-hand triode equal to the difference between the voltage drop across resistor R_3 and the voltage across the lower part of potentiometer R_4, goes down, too. Accordingly, the plate current of the right-hand triode builds up to a value at which the

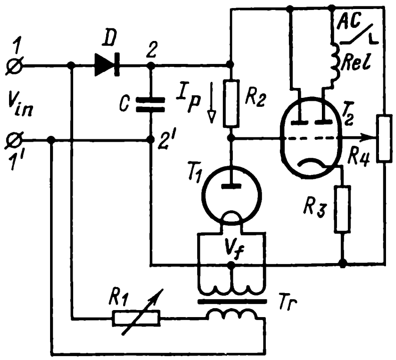


Fig. 21-3. A.c. vacuum-tube movable-contact relay

electromagnetic relay operates and its actuator circuit is completed.

When the input voltage, V_{in}, decreases, the anode current of the diode goes down, too, the diode anode voltage rises and the negative grid bias of the left-hand triode decreases. The plate current of the triode rises and causes the voltage drop across resistor R_3 to increase. The negative grid bias of the right-hand triode builds up, too, but the plate current decreases and causes the actuator circuit of relay *Rel* to open.

21-4. The D.C. RC-Network Timing Relay

In automatic circuits, use is widely made of electronic devices known as timing relays which introduce a definite time delay between the instant when the control circuit is opened or closed and the instant when the actuator (of final-control) circuit is completed.

The main time-determining element of a timing relay is an RC network. As stated in Sec. 18-5, the parameters of this network determine the rate at which capacitor C discharges aperiodically through resistor R connected between the cathode and grid of a vacuum tube. The circuit of a timing relay is shown in Fig. 21-4. The plate circuit of triode T, which contains the coil of an electromagnetic relay, *Rel*, is energized by the plate source, E_p. Contacts K are normally closed, and the negative voltage, V, which

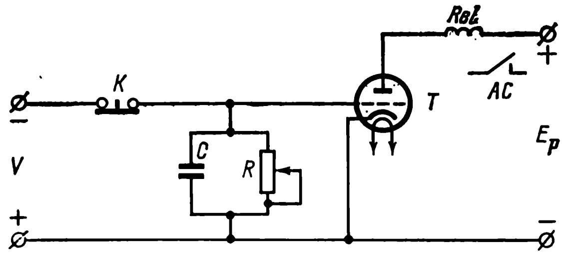


Fig. 21-4. Vacuum-tube timing relay

drives the tube to cut-off is applied between its grid and cathode. The same voltage is impressed upon capacitor C and resistor R. When contacts K open (the relay just begins to operate), capacitor C starts discharging through resistor R. The capacitor voltage, v_C, and the grid voltage of the tube decrease exponentially (Sec. 18-5), and the plate current of the tube rises in proportion. When the grid voltage reaches the value v_{C1} at which the tube current is equal to the operate current of the relay, the relay contacts in the actuator circuit make.

As it follows from Eq. (18-24), the voltage across the capacitor during discharge at an arbitrary time t is

$$v_C = V_C \exp(-t/\tau) = V_C \exp(-t/RC)$$

and, at the time t_1 when the relay operates,

$$V_{C1} = V_C \exp(-t_1/\tau)$$

whence, the time t_1 between application of power to the relay and operation of the relay is

$$t_1 = RC \log_e V_C/v_{C1}$$

The operate time t_1 of the relay is adjusted by varying the time constant $\tau = RC$ of the RC network, which is done by selecting an appropriate value for resistor R.

21-5. The A.C. RC-Network Timing Relay

Figure 21-5 shows the circuit of an a.c. timing relay. When contacts K are closed and a negative half-wave of the plate supply voltage induced into the secondary winding of transformer Tr is applied to the relay, capacitor C_1

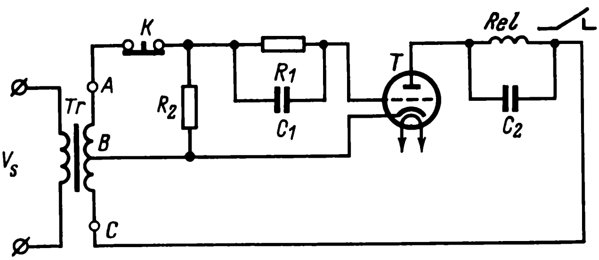


Fig. 21-5. A.c. RC-network timing relay

charges. The discharge of the capacitor through resistor R_1 produces a negative bias voltage which cuts off tube T. For the negative voltage at the tube grid to remain constant during the negative half-cycles of supply voltage, capacitor C_1 should have a sufficiently large capacitance, and the circuit time constant, $\tau = R_1C_1$, should be high as compared with the cycle of the alternating voltage. Then, the tube will be positively driven to cut-off, because the bias voltage will approximately be equal to the amplitude of the secondary voltage V_{AB} of the transformer.

Opening the contacts, K, disconnects the grid circuit from the secondary winding AB and connects it to the tube cathode via resistor R_2. Capacitor C_1 now discharges through resistor R_1, and the negative bias decreases exponentially. As a result, the tube turns on at a certain moment and its plate current rises to a value at which the relay operates. Naturally, the plate current can flow only during the positive half-cycles of plate voltage. To reduce the ripple current, the relay coil is shunted by capacitor C_2. The capacitor charges until the plate current reaches its maximum value and discharges through the relay coil when there is no plate current, thereby reducing the ripple in the coil. Relays which do not require d.c. power sources are simple in design and convenient in service.

21-6. The Transistor Timing Relay

Figure 21-6 shows the circuit of a basic transistor timing relay. When contacts K are closed, power source E charges capacitor C. When contacts K open (the relay starts to ope-

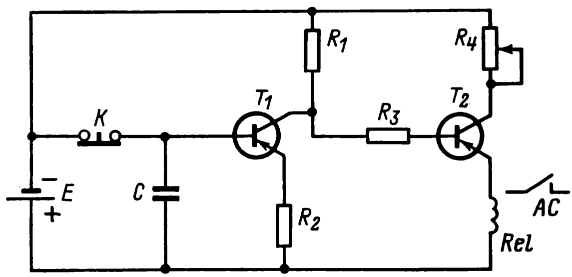


Fig. 21-6. Transistor timing relay

rate), the capacitor begins to discharge, the negative voltage at the base of the first transistor decreases, and the negative voltage at its collector builds up. Accordingly, the voltage at its collector builds up. Accordingly, the voltage at the base of the second transistor goes down, and in time t after the relay is energized, the emitter current reaches the value at which the electromagnetic relay, *Rel*, operates.

21-7. The Timing Relay Using a Glow-Discharge Thyratron

Figure 21-7 shows the circuit of a timing relay built around a glow-discharge thyratron. When contacts K are closed, a glow discharge occurs between the control electrode (grid) and cathode of the thyratron whose current is limited

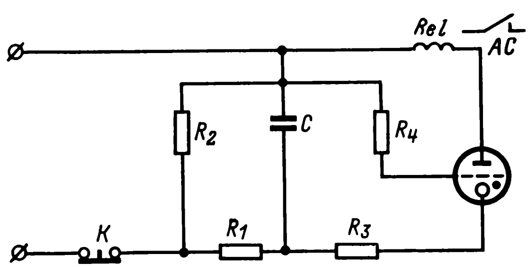


Fig. 21-7. Timing relay built around a glow-discharge thyratron

by resistor R_4. At the same time, capacitor C charges through resistor R_1. In accordance with Eq. (18-22), the voltage across the capacitor rises to the firing (or trigger) voltage of the thyratron, when a glow discharge strikes between the anode and cathode, and the capacitor begins to discharge via resistor R_3, the thyratron and the coil of relay *Rel*. The relay operates and closes its contacts in the actuator circuit. After the capacitor has discharged, the thyratron extinguishes.

Glow-discharge thyratrons make it possible to use less sensitive electromagnetic relays.

21-8. Photorelays

A photorelay is an automatic unit using a photo-electric device.

Figure 21-8*a* shows the circuit of a basic d.c. photorelay using a photoresistor, *PR*. The circuit contains a d.c. power source, *E*, and an electromagnetic relay connected in series with the photoresistor. The sensitivity and current of photoresistors are considerably greater than those of vacuum and gas-filled phototubes, so in many cases photorelays based on photoresistors do not use amplifiers, and their circuits are therefore extremely simple. When the photoresistor is not illuminated, its resistance is high, being about 1 megohm, and the current in the circuit is low. When illuminated, the photoresistor presents a low resistance and the current rises to a value at which the relay operates and completes the actuator circuit, *AC*.

The circuit shown in Fig. 21-8*b* is intended for operation from an a.c. power source. It differs from the previous circuit in that it has a crystal rectifier, *D*, connected in series with a shunt arm consisting of an electromagnetic relay, *Rel*, and a capacitor, *C*.

When a.c. power is applied and light strikes the photoresistor, one half-wave of current flows through its circuit during each positive half-cycle of voltage. During the negative half-cycles, no current flows through the photoresistor, but the relay coil remains energized because capacitor C charged during the positive half-cycle will now discharge through the coil.

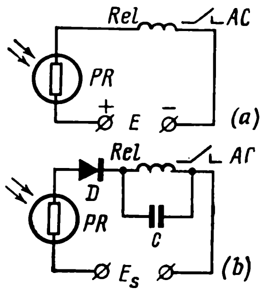


Fig. 21-8. Photorelay using a photoresistor

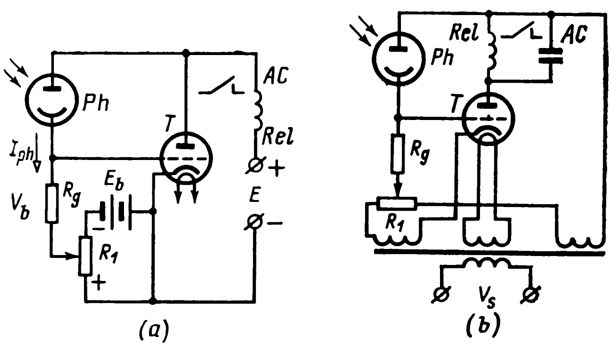


Fig. 21-9. Photorelays using vacuum-tube amplifiers
(*a*) front-effect relay; (*b*) photorelay operating on alternating current

Should the sensitivity of a photorelay based on a photoresistor prove insufficient, it will usually be ganged up with an electron amplifier.

According to the manner in which a photocell is connected in a relay circuit, one can obtain either a front-effect relay which operates when light strikes the photocell or a back-effect relay responding to darkness.

Figure 21-9*a* shows the circuit of a front-effect relay.

When no light strikes the photocell, *Ph*, its photocurrent I_{ph} is equal to zero and the tube, *T*, is turned off because a high negative grid bias, V_b, is applied from the grid bias source, E_b, to the tube grid. When light strikes the photocell, it develops a photocurrent I_{ph} which brings about

a positive voltage, $I_{ph}R_g$, across resistor R_g. This voltage reduces the negative grid voltage, and the tube is rendered conducting. Its plate current keeps rising until the illuminance reaches a certain value; at that instant the relay, *Rel*, operates and makes its contacts in the actuator circuit. When no light reaches the photocell, the relay contacts open.

If we interchange the photocell, *Ph*, and resistor, R_g, in the circuit of Fig. 21-9*a*, we shall obtain a back-effect relay. In this case, the relay contacts are closed in the absence of light because I_{ph} is zero and the negative grid bias circuit is opened. When the photocell is illuminated, the negative grid bias circuit is completed, the tube is driven to cut-off, the current in the relay becomes zero, and the actuator circuit is broken.

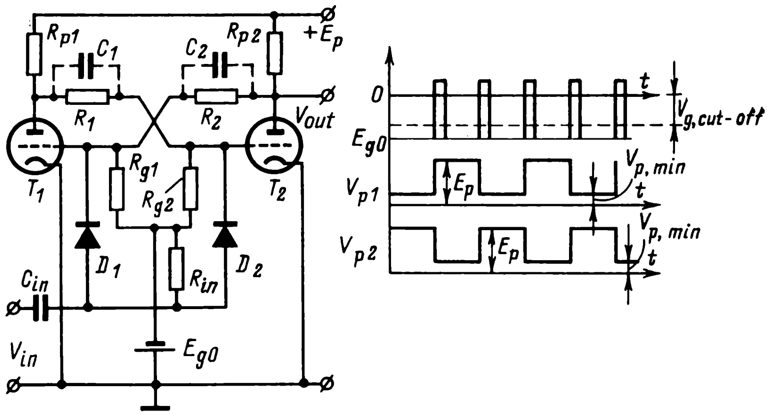
Figure 21-9*b* shows the circuit of a photorelay operating on alternating current. Current flows through tube T only during the half-cycles when the plate of the tube is positive. The initial value of grid bias determined by the wiper setting of potentiometer R_1 is negative and the total value is also controlled by the illuminance of the photocell. If enough light strikes the photocell, tube T is conducting during the positive half-cycles of applied voltage, the coil of the electromagnetic relay is energized and, if the coil current reaches the operate value, the relay contacts in the actuator circuit will make. During the negative half-cycles, the plate current is zero, and the current in the relay coil is maintained by the discharge current of capacitor C which was charged earlier.

Photorelays have found wide application in industrial electronics because they can be used to monitor and measure many quantities, such as illuminance, temperature and size of workpieces, transparency of materials, surface finish, number of workpieces on a production line, etc.

21-9. The Flip-Flop Circuit

A flip-flop circuit is a circuit having two stable states. Flip-flop circuits are often used as static electronic relays.

A flip-flop circuit built around vacuum tubes is shown in Fig. 21-10*a*.



(a) (b)

Fig. 21-10. (*a*) Vacuum-tube flip-flop circuit; (*b*) its timing diagram

This is a positive feedback flip-flop circuit in which the plate of the first tube, T_1, is connected via a voltage divider, R_1/R_{g2}, to the grid of the second tube, T_2, and the plate of tube T_2 is connected via a similar divider, R_2/R_{g1}, to the grid of tube T_1. Positive input pulses v_{in} are applied through diodes D_1 and D_2 to the grids of tubes T_1 and T_2. For the most part, flip-flop circuits are made symmetrical, that is, $R_{p1} = R_{p2}$, $R_{g1} = R_{g2}$, and tubes T_1 and T_2 are identical.

Owing to positive feedback, the state in which both tubes are conducting is unstable. The slightest asymmetry results in a transient which causes one tube to remain conducting and the other to be turned off. To prove, let the plate current, I_{p1}, of the first tube, T_1, exceed somewhat the plate current, I_{p2}, of the second tube, T_2. This will bring down the plate voltage V_{p1} of the first tube and, as a consequence, the grid voltage of the second tube, because the plate of the first is connected to the grid of the second. With time, the fall in the grid voltage of tube T_2 leads to a decrease in I_{p2} and, therefore, to an increase in V_{p2}. This rise in voltage is applied to the grid of T_1, and causes the plate current, I_{p1}, to build up still more. The transient process is completed

when tube T_2 is driven beyond cut-off and tube T_1 is conducting heavily. This state of the flip-flop circuit is stable, because it is maintained by the negative grid bias voltage supplied by the grid battery, E_{g0}, or an automatic bias RC-network.

Owing to positive feedback, the transients are complete in a fraction of a microsecond.

The flip-flop circuit remains stable until an external pulse causes it to change state. When a positive input pulse, v_{in}, is applied to the grid of cut-off tube T_2 (Fig. 21-10*b*), the latter starts conducting, its plate voltage goes down and is applied to the grid of conducting tube T_1, thereby causing its plate current, i_{p1}, to diminish. The fall in the current i_{p1} of tube T_1 forces its plate voltage and the grid voltage of tube T_2 to grow. The rise in the grid voltage of tube T_2 results in an increase in the plate current i_{p2}, and, therefore, in a fall in the plate voltage of tube T_2. This goes on until tube T_1 is driven to cut-off and tube T_2 turns on. This stable state of the flip-flop lasts until the next input trigger pulse is applied.

The flip-flop can be driven from one state to the other by negative pulses which should be applied to the grid of a conducting tube. In this case, the anodes of diodes D_1 and D_2 must be connected to the grid of the tube.

Capacitors C_1 and C_2 (Fig. 21-10*a*) connected in parallel (dotted lines) with resistors R_1 and R_2 are intended to increase the speed at which the flip-flop is driven from one state to the other, and also to store its previous state when trigger pulses are applied to both of its grids.

Flip-flops often use double triodes (for example, Soviet-made 6H2П, 6H15П or 6H8C triodes).

In addition to vacuum tubes, flip-flops can be built around transistors as shown in Fig. 21-11. In both circuitry and principle of operation, transistor flip-flops are similar to those using vacuum tubes.

If transistor T_1 is turned off, a high negative potential is applied from its collector through voltage divider R_1/R_{b2} to the base of T_2, so that, with suitably chosen values of R_1 and R_{b2}, the base of T_2 is a negative potential, transistor T_2 is conducting, and its collector voltage is close to zero. This means that the upper point of the divider (R_2/R_{b1})

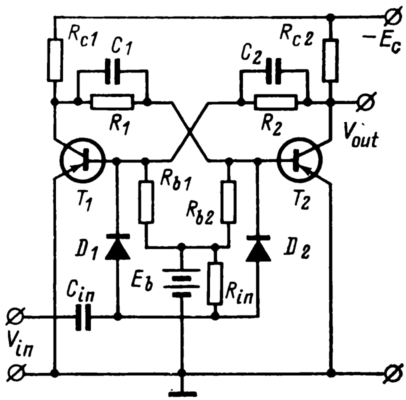


Fig. 21-11. Circuit of a transistor flip-flop

is at about zero potential and the lower point is at a positive potential equal to E_b. Therefore, the base of T_1 is positive relative to the emitter, and T_1 is turned off.

To transfer the triodes to the other state in which T_1 is on and T_2 is off, one should apply a positive pulse to the base of T_2. This causes the currents in the circuits to change in an avalanche manner so that transistor T_2 ceases conducting and remains in the off state until another trigger pulse is applied.

The major virtues of transistor flip-flops are as follows: they are small in size, light in weight and low in cost, draw little power, and are reliable in operation. Their main drawbacks are: low output voltage amplitude, $V_{out,\ \max} = 0.8$ to $0.9\ E_c$ ($E_c = 10$ to 15 V) and dependence on ambient temperature.

Flip-flops are widely used in automatic control, computers, and other equipment.

Chapter Twenty Two Fundamentals of Computers

22-1. General

The progress made by modern science and engineering has spurred the replacement of obsolete (mechanical, electromechanical and electric) computers by novel high-efficiency electronic computers. Apart from pure computational work, they control space craft and satellites, run nuclear reactors, are widely used in industrial automatic control and management systems, in defence, etc. In this chapter, we shall only discuss digital electronic computers which deal with mathematical variables in the form of numbers representing discrete values of physical quantities.

These computers can conventionally be divided into three classes:

1. General purpose computers designed to handle a wide variety of mathematical problems. For example, the ES-EVM and the BESM-6 solve complex scientific problems; the Ural-14 and the Ural-16 handle statistical work; and the NAIRI and the MIR which are small computers, deal with engineering problems.
2. Control computers, such as the Dnepr-1, the Dnepr-2 and the ASVT M-6000, designed to control production processes and run large-scale enterprises.
3. Special-purpose computers used for a narrow range of tasks. They are small in size and have limited potentialities.

22-2. Structure of a Digital Computer

If it is necessary to solve repetitive problems of the same type, for example, to extract the square root, we can work out a detailed and precise sequence of elementary opera-

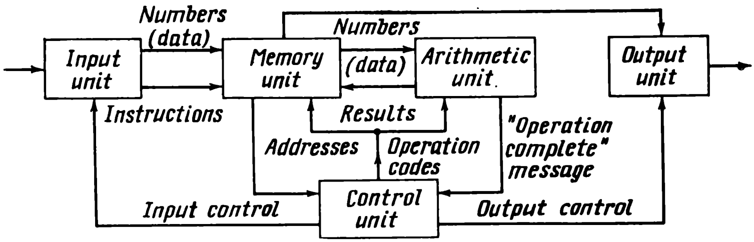


Fig. 22-1. Block diagram of an electronic digital computer

tions, by means of which any person could do the job in a routine manner, without comprehending the essence of the matter. Such a set of instructions, or program, is called the *algorithm*.

The word algorithm is derived from the Arabic "al-Khowarazmi", the man of Khiva, the surname of a ninth-century Uzbek mathematician who set forth the rules for working out such programs.

At the present-day level of electronic engineering, these elementary operations can accurately be perfomed by a digital computer at a speed unattainable for a human being, provided that it is supplied with a necessary program. A block diagram of such a digital computer is shown in Fig. 22-1.

The control unit (CU) of the computer sequentially interprets instructions specified by a program to direct arithmetic and logical operations. It is connected to an *input unit* (INPUT) and an *output unit* (OUTPUT). By means of these units the computer communicates with the outer world. They are not made integral with the computer (or processor), and are often called "peripherals".

The input unit feeds initial data into the computer by means of punched cards and tape, and the output unit presents final data as printed characters, curves, pattern on CRTs screens, or as holes in punched cards or tape. Thus, the input and output units convert data. As they use mechanical parts, their operating speed is low, being 600 punched cards per minute on the average, and 2000 cards per minute as a maximum. Inside the computer, the data exist as electrical signals, so the operating speed of the

computer is high, being over a million operations per second.

The control unit is connected to a *memory* or *storage unit* (SU) intended to store unitial data, programs, intermediate and final results, etc. In fact, there are two storage units, namely an external storage which has a large capacity but is slow in reading data "in" and "out", and an internal storage which only holds data used currently and not intended for long-time storage.

The *arithmetic unit* (AU) of the computer is intended to perform arithmetic and logical operations on numbers supplied by the storage.

As is seen, the control unit, storage and arithmetic unit are the major components of a computer, which organize the entire process of automatic computation.

22-3. Interaction of the Computer Units

The operation of a computer starts when a program is loaded by its input unit into its internal memory. The program presents instructions which specify all operations the computer is to perform, such as fetch number from such-and-such location, operate on the number, place the result in such-and-such location, stop the operation. All these instructions are handled in the form of coded electrical signals.

Every instruction consists of two parts: the operation part which specifies what should be done, and the address part defining the numbers of which the mathematical operation should be carried out. The operation to be performed is indicated by an operation code, for example, 01 may stand for addition, 02 for subtraction, 03 for multiplication, 04 for division, and so on. The second part of the instruction gives not the numbers themselves, but the locations from which they should be fetched. In the storage unit, each number is loaded into and retrieved from a separate storage location, or cell, which is asigned an address. For example, the instruction 01 0025 0030 0175 orders to add the number from cell 0025 to the number from cell 0030 and place the result into cell 0175 of the storage unit.

After the program has been loaded into the internal memory, the computation is carried out automatically.

As already noted, a computer is controlled by voltage or current pulses which are amplified, given a square waveform, and which, in certain combinations, may represent numbers. These pulses are formed in the control unit, and they are used to read numbers and instructions into and out of the storage unit. The number specified by the address part of an instruction is transferred from the storage to the arithmetic unit which carries out the arithmetic operation specified by the operation code. The result is transferred back to the storage unit at the address specified in the instruction. After an operation has been performed, the control unit receives a message from the arithmetic unit to that effect and passes on to the next instruction of the program.

22-4. The Binary Number System

In everyday calculations we generally use the *decimal positional number system*. Here, the symbols 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9, called digits, represent zero and the first nine whole numbers, or integers. The number 10 which is the radix (or base) of the system is designated by two digits, 1 and 0. The value (or weight) of each digit in this sequence depends on its position or place in the number. For example, in the decimal number 345.2, the fractional part is 2×10^{-1}, and the integer part consists of five unities (5×10^0), four tens (4×10^1) and three hundreds (3×10^2). The entire number can be written as follows

$$345.2 = 3 \times 10^2 + 4 \times 10^1 + 5 \times 10^0 + 2 \times 10^{-1}$$

To represent this number, the counter of any system should have ten stable states. However, it would take too much time and effort to build and use such a counter. So, instead of the usual decimal notation, electronic digital computers mostly use the *binary number system*, that is, a system with the base 2. This system requires counters with only two stable states, known as scale-of-two or binary counters because this system uses only two digits, 0 and 1. To write the decimal 2 in binary notation, we place a 0 in

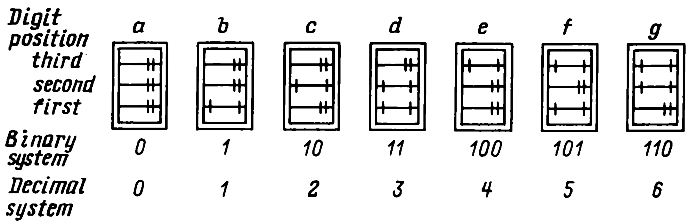


Fig. 22-2. Binary numbers counted on an abacus

the least-significant position and a 1 in the next higher position. So, a 1 in any higher position is equal to two unities in the next lower position.

This can easily be illustrated with the help of an abacus having only two balls on each wire (Fig. 22-2). In the "a" position, the counter reads zero in both the decimal and binary systems. In the "b" position, it reads 1 in both systems. In the "c" position, the decimal number is 2 and the binary one is 10 (which should be read as one-zero). For the decimal 3, the binary equivalent is 11 (one-one), and so on. For example, the binary number 1011.1 corresponds to a decimal number which has one half (1×2^{-1}) in the fractional part, and one 1 (1×2^0), one 2 (1×2^1), no 4's (0×2^2) and one 8 (1×2^3) in the integer part. That is, $1011.1_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} = 8 + 0 + 2 + 1 + 0.5 = 11.5_{10}$; here the subscripts *2* and *10* denote base *2* and *10*, respectively.

Thus, in a binary computer, the counter should have for each bit (binary digit) a device operating as a bistable circuit, that is, one with two stable states, ON — OFF. This function is usually performed by flip-flops (see Sec. 21-9).

The flip-flop can be set to these states by ON/OFF control (clock) pulses, by the presence or absence of voltage, by a magnetized or a demagnetized area on magnetic tape, etc. For example, the binary number 10111 can be transmitted by the voltage pulses shown in Fig. 22-3.

The main drawback of the binary system is that the counter needs a great number of digit positions as compared with the decimal system, but this is made up for by the simplicity of the computer design. In addition to the binary

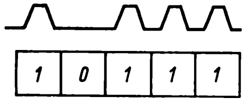


Fig. 22-3. Binary number encoded by a train of pulses

representation, electronic digital computers use the octal and hexadecimal number systems, the former for program writing and the latter for readout, because an octal or a hexadecimal number is much shorter than the respective binary number. These systems will be discussed in greater detail later.

22-5. Arithmetic Operations on Binary Numbers

An important merit of the binary system is that the arithmetic operations on binary numbers are performed as with decimals. For this purpose, we can use the tables given below.

(a) Addition

Example 22-1. Add together two numbers. Their decimal representation is on the left and the binary representation on the right

$$
\begin{array}{rr@{\qquad}rl}
 & 1 & 11 & \text{— carry} \\
 & 25 & 11001 & \text{— augend} \\
+ & & + & \\
 & 19 & 10011 & \text{— addend} \\
\hline
 & 44 & 101100 & \text{— sum}
\end{array}
$$

In the decimal numbers system, the addition of 5 and 9 in the units position gives 14. The "4" is left in that position and the "1" (which actually is "one ten") is moved to the tens position and added to the number of tens already present: $2 + 1 + 1 = 4$.

A similar procedure applies to the binary system. On adding two unities in the first position, we leave a 0 in that position and move a 1 (which actually is "two unities") to the next more significant, or second, position. On adding

two 1's in the second position, we again leave a 0 in that position and move a 1 (which now is "four unities") to the third position and so on.

(b) Subtraction

Example 22-2

$$\begin{array}{rrl} & \dot{2}5 & \quad 1\dot{1}0\dot{0}1 \text{ — minuend} \\ - & 19 & \quad 10011 \text{ — subtrachend} \\ \hline & 06 & \quad 00110 \text{ — difference} \end{array}$$

In the decimal system, we have to borrow 1 from the tens order, and the difference is equal to six. In the binary system, we can, if necessary, borrow unity from the next higher position. The unity borrowed from the second position is obviously equal to two unities in the first, that borrowed from the third position is equal to four unities in the first position, etc. In our example, the difference in the first position is zero. In the second position, we subtract its unity from that borrowed in the third position and equal to two unities in the second position; this leaves a difference of 1. Similarly, the difference in the third position is 1; in the 4th and 5th positions it is zero.

In computers, subtraction is usually replaced by addition, in which the subtrachend is written in its complement form. This means that all its unities are replaced by zeroes, and all its zeroes by unities. This is illustrated by Example 22-3.

Example 22-3.

$$\begin{array}{rr} + & 25 \\ & 81 \\ \hline & (1)\,06 \end{array} \qquad \begin{array}{rr} + & 11001 \\ & 01100 \\ \hline & 100101 \\ & \longrightarrow 1 \\ \hline & 110 \end{array}$$

In the decimal representation, this is done as follows: the minuend 25 is added to the complement of the subtrachend, that is, the number which complements it to 100

(100 — 19 = 81), and the unity in the most significant (hundreds) order in the sum is discarded. As in conventional subtraction, the difference is six.

In binary notation, 11001 is added to 10011 written in the complement form, that is, as 01100. Then, the unity in the most significant position is transferred to the least significant position and added to the bit it contains. The result is the same as in the case of subtraction. This is automatically carried out by the arithmetic unit of a computer.

(c) Multiplication

In this case, use is made of the addition table.

Example 22-4. Multiply 4 (decimal) by 5 (decimal), or, in the binary system, 100 by 101:

$$
\begin{array}{r}
\times\begin{array}{r}100\\101\end{array}\\
\hline
100\\
100\\
\hline
10100
\end{array}
$$

As is seen, if a bit of the multiplier is 1, the multiplicand is shifted as many places as may be necessary, and, if a bit of the multiplier is 0, the multiplicand is shifted left one place. As is seen, the multiplication of binary numbers is accomplished by the combined operations of shifting the multiplicand one place and addition, which are carried out by the arithmetic unit.

(d) Division

Division is done by the repeated subtraction of the divisor from the dividend and complementation of the remainder from the right. Here, use is made of the multiplication and subtraction tables.

Example 22-5.

$$
\begin{array}{rl}
111000 & : \; 1000 \\
-\underline{1000} & \;\;\; \overline{111} \\
-1100 & \\
\underline{1000} & \\
-1000 & \\
\underline{1000} & \\
0000 &
\end{array}
$$

or, in the decimal number system, $56 : 8 = 7$.

Since multiplication is replaced by repeated addition, division by repeated subtraction, and subtraction can be replaced by addition, all arithmetic operations in a computer reduce to addition.

22-6. The Operating Principle of Some Computer Elements

A present-day electronic digital computer consists of simple logic elements designed on the basis of mathematical logic (known as Boolean algebra) which was introduced by English mathematician George Boole (1815-1864). Some of these elements implementing the main logical functions, namely *NOT*, *AND* and *OR* gates, are considered below.

(a) The NOT Gate

The circuit of the *NOT* gate is shown in Fig. 22-4. Its input, A, accepts a signal from the collector load of a previous circuit which is not shown in the figure. The signal varies from $-V_0$, corresponding to a logical 0, to zero, corresponding to a logical 1. The output, P, delivers an inverted input signal $P = \bar{A}$, which should be read as follows: P is not A. The transistor operates as a switch, that is, i is periodically turned ON and OFF (see Sec. 19-11).

Let, initially, a logical 0, that is, a negative voltage, $-V_0$, be applied to input A. As a result, the necessary

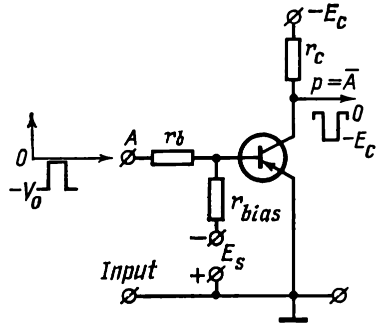


Fig. 22-4. Logic *NOT* gate

potential difference is established between the base and emitter, the transistor is rendered conducting and practically all of the collector voltage ($-E_c$) is dropped across resistor r_c; the potential at the input becomes close to zero.

When a logical 1, that is, a zero voltage, is applied to the input, the base potential increases, the transistor is turned off. The current in the transistor becomes zero and the output potential is practically equal to $-E_c$.

(b) The AND Gate

Figure 22-5 shows a two-input diode-resistor *AND* gate. This circuit realizes the logical operation $P = A \wedge B$ which is read as follows: the operation P cannot be performed unless the operations A and B are carried out simultaneously. The number of inputs in the gate may be more than two. The forward diode resistance in this circuit is not over 100 to 200 ohms, and the reverse resistance is about several megohms. The value of resistor r_{and} is several kilohms.

If, initially, signals $-V_0$ (logical 0's) are applied to both inputs A and B, one or both diodes conduct current under the action of $+E_{and}$. All of the voltage is dropped across resistor R_{and}, and a low potential $-V_0$ is passed on to output P. If logical 1's, that is voltages, are simultaneously applied to the inputs, both diodes cease conducting, the current in the resistor goes down, and a zero potential,

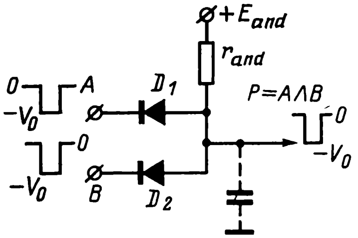


Fig. 22-5. Logic *AND* gate

that is, a logical 1, appears at the output. When at least one input accepts a logical 0, the respective diode will be driven to conduction, and the voltage at the output will drop to $-V_0$ (a logical 0).

The timing diagram of a three-input *AND* gate is shown in Fig. 22-6. A signal appears at output P only when logical 1's are applied to all inputs simultaneously. So this gate realizes the logical function $P = A \wedge B \wedge C$.

Figure 22-7 shows a two-input *AND* gate intended to gate or delay a signal, and having no separate power source E_{and}. The circuit operates as a switch, because a diode has two states, one in which it conducts current (the *ON* state) and the other in which it conducts no current (the *OFF* state).

If a square positive pulse is applied to the upper point of input A, a current will flow through the diode from the anode to the cathode and via the control (or clock) input circuits (not shown in the figure) to ground. The resistance of resistor r is high and that of the diode is negligible, so all of the applied voltage is dropped across resistor r. The potential at output B does not differ from the ground potential.

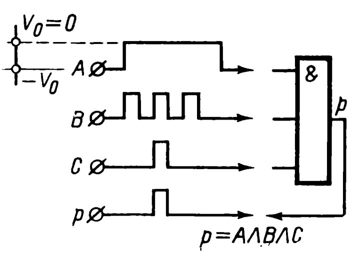


Fig. 22-6. Pulses in a three-input *AND* gate

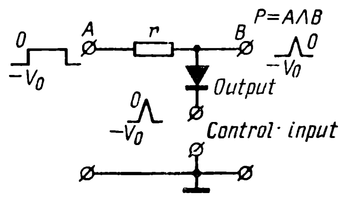


Fig. 22-7. Explaining operation of a diode

Thus, the output delivers a logical zero—no pulse has passed through. When positive pulses are simultaneously applied to input A and clock input, the current drops to zero, the voltage drop across resistor r is also zero, the potential at point B becomes equal to the potential at point A, so the output delivers a logical 1.

(c) The OR Gate

Figure 22-8 shows a diode-resistor *OR* gate. It implements the logical operation $P = A \vee B$ which is read as follows: the operation P can be performed when a signal is applied to input A, or input B, or to both inputs simultaneously.

To minimize distortion, the circuit parameters are chosen such that $R_{or} \gg R_{d,f}$ and $|-E_{or}| > |-V_0|$.

Initially, both diodes are turned off because $-V_0 < -E_{or}$. The voltage at the output is $-E_{or}$. When a positive signal (a logical 1) is applied, for example, to input A, diode D_1 is rendered conducting, the current in resistor r_{or} rises, and the voltage drop across the resistor is practically equal to $-E_{or}$. Now, the output voltage corresponds to a logical 1. When the signals are simultaneously applied to inputs A and B, the current in and voltage drop across

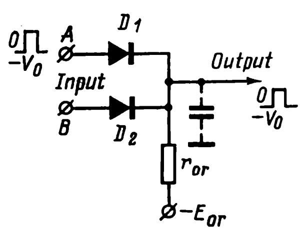


Fig. 22-8. Logic *OR* gate

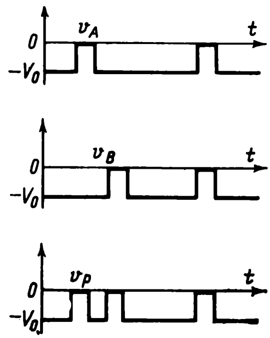


Fig. 22-9. Pulses in an *OR* gate

resistor r_{or} do not change because the forward resistance of the diodes is low as compared with r_{or}. So, the signal amplitude at the output remains unchanged. The operation of the gate is illustrated in Fig. 22-9.

(d) The Diode Network with One Control Input

Figure 22-10 shows a diode network with one control input. The pulses from the inputs may pass on to the out-

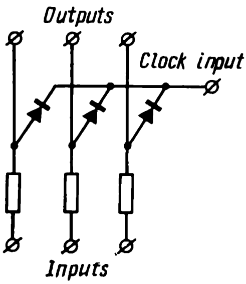


Fig. 22-10. Diode network with a common control (clock) input

puts only when the control input accepts a pulse which cuts off the diodes. In the absence of such a pulse, the positive input signals cause a current flow through the diodes, and, owing to the high voltage drop across the resistors, the anode and output potentials fall close to zero.

(e) The Shifter

The circuit of a one-position shifter is shown in Fig. 22-11. As already noted, in binary multiplication by a 1 bit system, the multiplicand is written into the position of that bit, and, in binary multiplication by zero, the same number is written shifted one place. This operation is carried out by the shifter.

If a high potential exists only at the digit inputs and absent at buses *I* and *II*, currents flow only through diodes

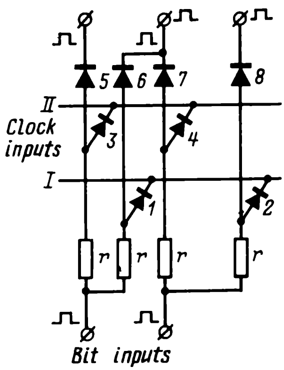


Fig. 22-11. Circuit of a one-position shifter

1, *2*, *3*, and *4* and output potentials are very low (no signals). When a positive pulse is applied to bus *I*, diodes *1* and *2* are turned off, and the pulses pass through diodes *6* and *8* onto the middle and right-hand outputs. Diode *7* is cut off by the high potential at the left-hand input. If a pulse is applied to bus *II*, diodes *3* and *4* cease conducting. The pulses pass through diodes *5* and *7* onto the left-hand and middle outputs (that is, they are shifted to the left), and diode *6* is cut off by a high potential.

22-7. The Operating Principle of the Binary Counter

The flip-flop considered in Sec. 21-9 can store one bit of a binary number and respond to the presence (1) or absence (0) of a digit in the respective position. Thus, as many flip-flops must be used as there are bits in the number.

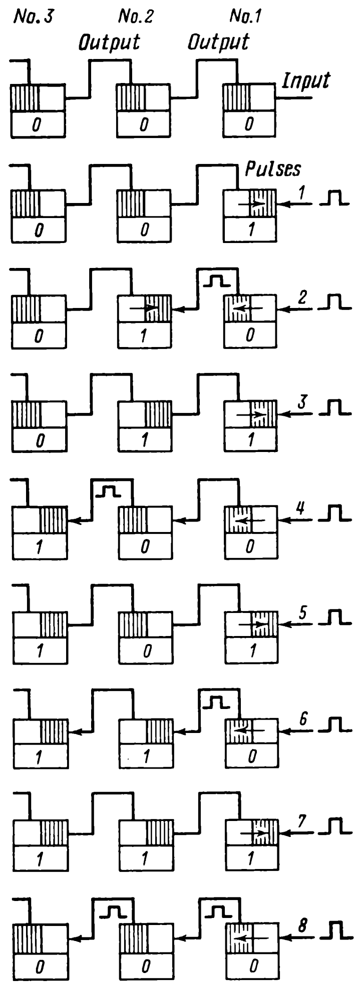


Fig. 22-12. Explaining operation of a binary counter

A chain of flip-flops intended to store one binary number is called a *register*.

A *counter* consists of series-connected flip-flops and is used to count the number of pulses applied to its input. Figure 22-12 illustrates the operation of a three-bit counter.

Let flip-flops No. 1, No. 2 and No. 3 be in such a state that their left-hand (shaded) parts are conducting and the right-hand (unshaded) parts are not conducting. This state is assumed to be 0, then the number stored by the flip-flops will be 000.

Now, let us apply a sequence of uniformly spaced pulses to the input of flip-flop No. 1. The first pulse drives flip-flop No. 1 to the "1" state, as shown by the left-to-right arrow in the second line of Fig. 22-12. As a result, the counter stores the number 001. The second pulse drives flip-flop No. 1 back to the 0 state. In doing so, the first flip-flop additionally transfers an output pulse to flip-flop No. 2, which causes it to change state from left to right. Now, the counter stores the number 010. When the third pulse is applied to the input of flip-flop No. 1, another 1 is read into the counter, so it stores the number 011. The fourth pulse also reads in a 1, but since the first flip-flop has changed state from right to left, it transfers an output pulse to flip-flop No. 2. The latter has changed state from right to left, so it delivers an output pulse in turn to flip-flop No. 3. Now the counter stores 100. This goes on until the eighth pulse sets the counter back to the initial position. As is seen, the three-bit counter can count decimal numbers from 0 to 7.

22-8. The Operating Principle of the Adder in the Arithmetic Unit

Let us add together two binary numbers $A = 1110$ and $B = 1101$. The addition is performed sequentially bit by bit as in the decimal system:

$$\begin{array}{l} C = 11 \\ A = 01110 \\ \underline{B = 01101} \\ S = 11011 \end{array}$$

The sum of the least significant bits is $0 + 1 = 1$. The same applies to the next bits: $1 + 0 = 1$. The sum of the third bits, $1 + 1$, is two unities, which is more than a bit position can hold. Accordingly, we leave 0 in the third

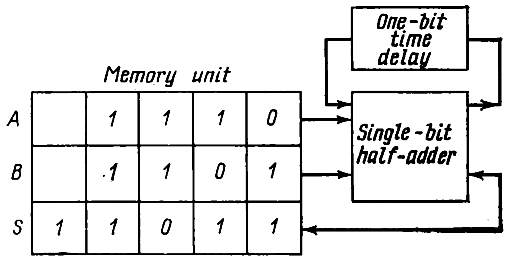


Fig. 22-13. Explaining operation of a full adder

position and are said to "carry" 1 to the next higher (fourth) position. This 1 is written in the C (carry) line above the number A. The sum of the fourth bits plus the carry, $1 + + 1 + 1$, is three unities—one unity is written in the S (sum) line, and the two remaining unities generate a carry, so a 1 is carried to the fifth position where $S = C + A + + B = 1 + 0 + 0 = 1$.

It is obvious that a device for serial addition as performed above, called the full adder should have two number inputs A and B, a carry input (from a lower-order position) C, a sum output S and a carry output (to a higher-order position) C. The block diagram of a multibit full adder using a memory unit, a single-bit half-adder and a delay circuit which stores the carry for one digit time, is shown in Fig. 22-13.

The circuit of the full adder (Fig. 22-14) consists of elementary logic gates, namely *AND*, *OR* and *NOT*, considered above. For simplicity, all grounding, pulse-forming and amplifier circuits are not shown.

Because all *AND* gates are connected to the positive terminal of the power source E, a current flows through their resistors r and the internal resistance of signal sources. The negative terminal of the signal sources is grounded. The anode potentials of the diodes in these circuits are low because a considerable part of voltage drops across resistors r. High potential pulses appear at the anodes only when positive voltage pulses high enough to cut off the diodes are applied to all cathodes. It is only then that the *AND* gate can pass on the positive pulses.

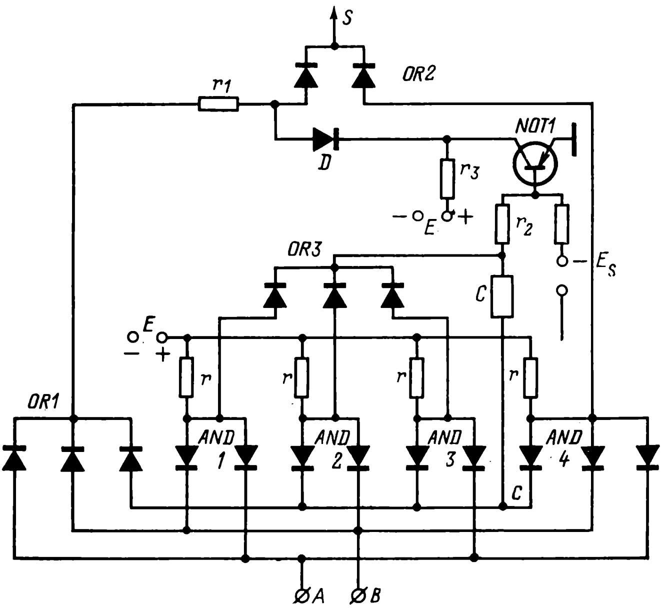


Fig. 22-14. Circuit of a full adder

A positive pulse can pass through the *OR* gates if it is applied to the anode of at least one diode. The *NOT1* gate is turned off by a negative potential at the transistor base. Now, the potential at its collector is high because no current flows in resistor r_3. This potential cuts off diode D, and it is only in this state that positive pulse can pass through resistor r_1 onto output S. The circuit operates as follows. When a positive voltage pulse is applied to the base of the transistor in the *NOT1* circuit, the transistor is rendered conducting, the potential at the collector goes down and diode D turns on. In this case, when the pulse is applied from the *OR1* gate, the anode potential of diode D is low because the current flows through the *NOT1* gate. The potential at output S is also low.

Now we can discuss the addition of the same two numbers. Before the operation, the flip-flops of the sum register S in the memory device (Fig. 22-13) are set to zero. The control

unit of the computer generates clock pulses which control the circuit. During each clock pulse, the bits in the same position of the augend and addend are added together.

During the first clock pulse, the bits in the least significant position, $A = 0$ and $B = 1$, are fetched from the memory device. In other words, a positive pulse is applied to input B of the adder (Fig. 22-14), and no pulse is applied to input A. Now the pulse can pass through the *OR1* gate, resistor r_1, and *OR2* gate onto output S, and drive the lowest-order flip-flop of the sum register S in the memory device to the "1" state.

During the second clock pulse, the digits $A = 1$ and $B = 0$ are fetched. A pulse applied to input A (Fig. 22-14) passes through the *OR1* and *OR2* gates onto output S and drives the second-bit flip-flop in the memory device to the "1" state (Fig. 22-13).

The third clock pulse causes the digits $A = 1$ and $B = 1$ to be fetched from the memory device, so positive pulses are applied to inputs A and B of the adder. Now the *AND1* gate turns off and applies a positive pulse to the *OR3* circuit. The pulse passes on to delay circuit C and the base of the transistor in the *NOT1* circuit, and the transistor turns on. Diode D is rendered conducting, and pulses A and B pass through the *OR1* gate, diode D and the transistor of the *NOT1* gate. The anode potential of diode D becomes low, and no signal is passed through the *OR2* gate on to output S. The third-bit flip-flop in the memory device remains in the "0" state.

The voltage pulse is delayed in the delay circuit until the bits in the fourth position are added together. This is not unlike a person who keeps a carry in his mind. The structure of the delay circuit will be discussed later.

During the fourth clock pulse, the augend and addend pulses A and B and the carry pulse C from the delay circuit cut off the *AND2*, *AND3* and *AND4* gates, and these apply a second pulse via the *OR3* gate to the delay circuit and the *NOT1* gate. As already noted, the *OR1* gate and diode D would not pass a pulse on to output S in such a case. Since, however, the *AND4* gate is turned off by the high potentials at inputs A and B and the first pulse C from

the delay circuit, its anode potentials go high, and a high voltage pulse is passed through the *OR2* gate and the output to the memory device. As a result, the fourth-bit flip-flop is driven to the "1" state.

During the fifth clock pulse, pulses A and B are absent, but the carry pulse from the previous (lower-order) position is allowed to pass through the delay circuit and *OR1* and *OR2* gates and output S to the fifth-bit stage of the adder, and the flip-flop of that stage stores 1. Thus, the sum is $1110 + 1101 = 11011$ or, in the decimal system, $14 + 13 = 27$.

22-9. Delay Lines

Delay lines are memory devices that can store data for a very short time (microseconds), after which the data are destroyed. These lines can be coaxial cables, waveguides, or series-connected LC-filters. Figure 22-15 shows a sketch of an acoustic delay circuit.

Two quartz plates, *3*, are inserted with the help of rubber rings, *4*, in a steel pipe, *1*, filled with mercury, *2*. A quartz plate changes its volume when moved in a varying electric field. A voltage pulse passing from one plate to the other causes mechanical oscillations in the first quartz plate and the mercury filling the pipe. These oscillations are transmitted at a certain speed to the other quartz plate at the output. The output quartz plate converts these mechanical oscillations into electrical oscillations. The speed at which the mechanical oscillations are transmitted through the mercury is incomparably lower than the speed of propagation for electrical oscillation, so the electrical pulse is transmitted with a delay. The delay time is usually a few microseconds. It can be varied by varying the pipe length.

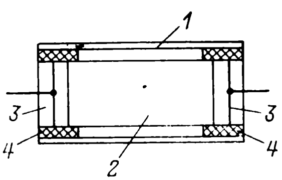


Fig. 22-15. Construction of an acoustic delay line

22-10. Memory Units

In every-day calculations we generally use our own memory. Some information has only to be kept in mind for a short time. For example, in the case of addition, we have to remember that a unity should be carried to the next higher position. Some information, for example, the multiplication table, should be kept in memory for a long time, though there is no need to memorize the values of sines, cosines or logarithms because they can be taken from handbooks.

We have already discussed some of the memory devices used in a computer, such as a flip-flop register and delay lines. However, a flip-flop memory device for a great number of multibit numbers would be too bulky.

The memory units of a computer are divided into two basic groups, namely *internal memory* (or *storage*) and *external memory* (or *storage*). Internal memory communicates directly with the arithmetic unit and determines the operating speed of a computer. It has a comparatively small capacity—tens of thousands of numbers at most—but it can read them in and out quickly (in a fraction of a microsecond). External memory does not communicate directly with the arithmetic unit, rather it does so through the internal memory and provides backup storage. It can store as many as billions of digits, but reads them in and out in blocks. The access time of external storage is long, being up to tens of milliseconds.

Internal memory can be constructed with toroidal cores fabricated from a ceramic ferrite material which has a rectangular hysteresis loop. Figure 22-16*a* shows a ferrite core with two windings; its magnetization curve is given in Fig. 22-16*b*. The state of positive induction $+B_0$ is taken as a logical 1; the state of negative induction $-B_0$ is taken as a logical 0.

Assume that there is no current flowing in the write winding w_1 and the sense winding w_2, as they are called. In the circumstances, the remanent induction is characterized by point, *2*, that is, $-B_0$, and a binary 0 is said to be written in the core. When a positive pulse applied to the write winding w_1 is such that the magnetic field inten-

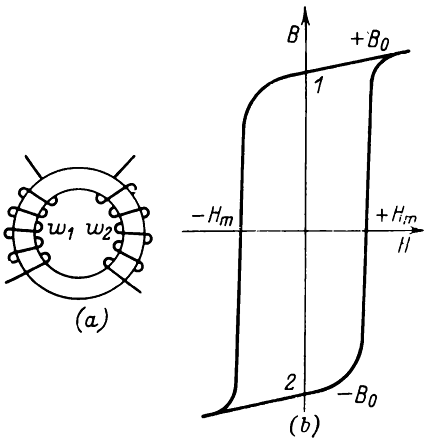


Fig. 22-16. Explaining operation of a magnetic core as a memory cell

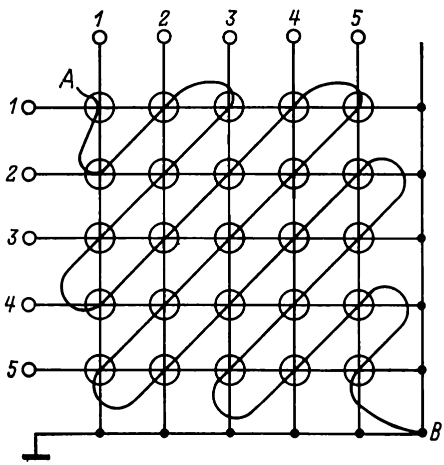


Fig. 22-17. Magnetic-core memory plane

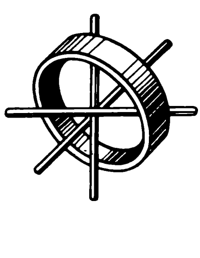


Fig. 22-18. Magnetic core in a memory plane

sity exceeds $+H_{m0}$ for an instant, the remanent induction after the pulse ceases will be $+B_0$, and the core will now store a binary 1. The next positive pulse will leave the remanent induction the same, $+B_0$, but a negative pulse will magnetize the core in the reverse sense, so a 0 will be written in.

As is seen, the core behaves like a flip-flop, and numbers can be written in and read out of a set of cores in less than 1 μs.

If the applied pulse reverses the polarity of the core, an emf is induced in the sense winding; if it does not, practically no emf appears there. In this way we can sense in which state the core was.

Such ferrite cores are strung on wires to form a *magnetic core memory plane* (also called a core plane or a memory plane), each using a hundred thousand cores. A sketch of a memory plane is shown in Fig. 22-17. Instead of windings, two wires run through each core as shown in Fig. 22-18. The horizontal and vertical wires, called the *write wires*, are used to write (read in) data, and the diagonal wire AB, called the *sense wire*, serves to read them out. Each horizontal row stores one binary number, and as many vertical rows are provided as there are bits in the binary numbers to be stored. The initial state of all cores is zero.

Assume that the binary number 1101 should be written in the second row (Fig. 22-17). Then half the current needed to magnetize a core, that is, a current producing a field intensity of $+H_m/2$, is applied each to the second horizontal row and to the second, third and fifth vertical rows. These half currents establish a magnetic field intensity H_m only where they coincide, that is, in the cores at the intersections of these wires. The state of magnetization of these cores is changed and they store each a 1. The magnetic field intensity in the cores of the first and fourth vertical rows is $H_m/2$ and their state of magnetization is not affected. As a result, the binary number 1101 is written in. In this case, the voltage pulse generated in the diagonal sense wire is ignored owing to the arrangement of the external circuits of the memory plane.

The numbers stored in other horizontal rows of the memory plane remain the same because in writing the number

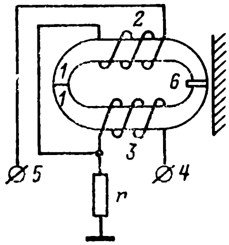


Fig. 22-19. Write and read magnetic head

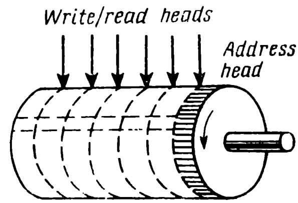


Fig. 22-20. Magnetic drum

1101 the half-currents in the vertical wires could not establish in them a magnetic field intensity greater than $+H_m/2$.

To read out a number is to sense in which state, 1 or 0, the cores are involved. If no emf is generated in the sense wire when a current pulse is applied, the core stores a 1. If the pulse generates emf, the core stores a 0. When reading out a number, the half-currents developing a magnetic field intensity of $H_m/2$ are applied to the sense wire and one of the horizontal rows corresponding to the address of the number being read out. In this case, the half-currents add together and reverse the polarity of only the cores in that row, and voltage pulses are generated in the respective vertical wires (representing the bits of the number being read out). During readout by reversal of polarity, the 1's stored in the cores are erased. However, it is usually desirable that the number read out should remain in the memory plane. Accordingly, it has a capability to restore a 1 after it has been read out. Memory planes can store each 100,000 binary digits or more and can store them as long as necessary.

Electronic digital computers widely use magnetic memory devices other than cores, namely magnetic tape, drums, discs. They serve as external and buffer memory (buffer memory is one between external and internal memory). Basically they operate as follows. A moving surface (a tape, a drum or a disc) coated with a magnetic material having a high remanent induction passes by a head which is essentially an electromagnet. Current pulses corresponding to the bits of the number energize the electromagnet, and this

magnetizes a spot on the magnetic surface, thereby producing a record of the number.

As a rule, numbers are written in and read out by the same head constructed as shown in Fig. 22-19. The head core, *1*, is made of a ferromagnetic sheet material having a low remanent induction and carries two windings, *2* and *3*. A unity is written in when a pulse is applied to terminal *4*, and a zero when a pulse is applied to terminal *5*. Numbers are read out at terminal *4*. The air gap, *6*, between the poles of the core is formed by a brass-foil spacer several micrometers thick.

The number of heads should be equal to that of bits in the number being recorded, as is shown in Fig. 22-20 for a magnetic drum. Here, an aluminium drum is coated with a layer of a magnetic material; heads are located above its surface along the generating line of the cylinder at the points marked by arrows. The drum has 5 to 8 heads per centimetre of its generating line. A number is recorded also along its line. For example, to record the binary number 1011, the pulses shown in Fig. 22-21*a* should be applied to the heads. As a result, magnetized spots (dipoles) appear on the drum surface (Fig. 22-21*b*). During readout, a magnetized spot on the rotating drum passes by the air gap of the head and generates emf pulses in the read winding. These pulses are shown in Fig. 22-21*c*. They are amplified and converted into square pulses which are transferred through the internal memory to the arithmetic unit.

In recording, as many tracks of bits are formed around the circumference of the drum as there are heads. The recording density is 1 to 3 dipoles per millimetre. The capacity of magnetic drums runs up to 1.5-2 million bits. One track and one head (Fig. 22-20) read the addresses of recorded bits. This head is connected to a bit counter. The drum is constantly rotating, so numbers can only be read out in turn. The speed of a drum driven by an electric motor is $n = 6000$ to 12,000 rpm, and the time required to locate the necessary bit is a split second.

Magnetic tape recording is carried out in about the same manner. Magnetic tape is made up of an elastic base coated with a layer of varnish mixed with a ferromagnetic powder. Heads are placed across and tracks run along the tape

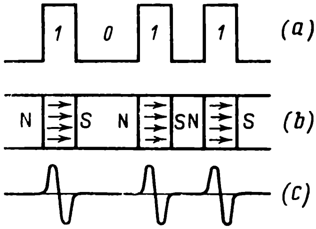


Fig. 22-21. Magnetic dipoles in recording a binary number

The readout speed depends on the speed at which the tape can be driven. Magnetic tape is used in auxiliary or secondary low-speed memory devices whose capacity is practically unlimited.

22-11. Input and Output Units

At first, a program for a computer is compiled by a programmer in digital form on paper. The computer cannot accept the program as it stands. The data encoded as digits in the program should be converted into a form acceptable for processing in the computer. Most often, input media for digital computers are punched cards made of dense elastic paper or punched tape. The holes representing information are made in cards and paper tape on devices called card punches or tape punches. Cards and tape are of a standard size; tape is stored in reels.

As electronic digital computers use the binary numbers system, one could think that a binary 1 may be represented by a hole and a binary 0, by no hole. However, programs are written in the decimal system, and the instructions and numbers (data) loaded into a computer may run into several thousands. Hence, it would be necessary to convert decimals into binaries and punch holes for a great number of multibit binaries, which would be to no purpose. Therefore, a punch first converts a decimal number into a binary-coded decimal and makes holes (punchings) according to this binary-coded decimal (BCD) representation. This is done as follows.

Any decimal digit can be represented by a group of four bits known as a *tetrad*: 0 = 0000, 1 = 0001, 2 = 0010,

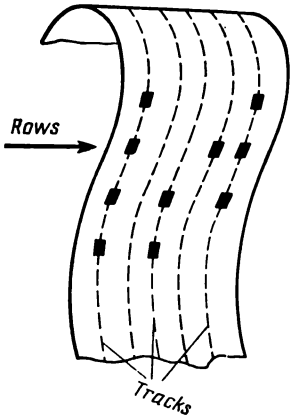


Fig. 22-22. Decimal number in BCD representation

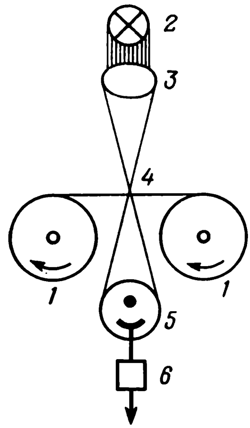


Fig. 22-23. Input unit of an electronic computer

$3 = 0011$, $4 = 0100$, $5 = 0101$, $6 = 0110$, $7 = 0111$, $8 = 1000$ and $9 = 1001$.

Then, making a hole for a binary 1 and no hole for a binary 0 on a card or tape, we can represent any number. Figure 22-22 shows the binary-coded decimal 1354 punched on tape. On four tracks or levels, from the right to left, the punch makes holes corresponding to the tetrads of the decimals 1, 3, 5 and 4. In the left (fifth) track, additional holes necessary for the readout are always punched. This track is known as a *clock track*. Thus, the punch has performed the first input operation—it has converted the decimal number into the binary-coded decimal and made the holes.

Now the punched tape which is a kind of a memory device is sent to the input unit of a computer where the holes representing numbers are converted into electrical signals. A simplified structure of an input unit is shown in Fig. 22-23. The tape transport has two reels, *1*, and the tape is transported from one reel to the other. Light from a lamp, *2*, passes through a lens, *3*, and a hole, *4*, in the nontransparent tape and strikes a photocell, *5*, which generates a low current pulse. The pulse flows through an amplifier, *6*, and appears at the memory unit which registers a binary *1*. When there

is no hole in the tape, a binary 0 is registered in the memory. Each track has a photocell and amplifier of its own.

However, if we read back the tape in tetrads, we would get the number 0001 0011 0101 0100 and not the number 1354_{10} in the binary system. So the computer is provided with a program according to which it converts the binary-coded decimal to the binary number and only after that it is registered in the memory.

To obtain the decimal 1354 we should add together 1×10^3, 3×10^2, 5×10^1 and 4×10^0. In a like manner, the computer receiving the tetrads one after another multiplies them by the position multipliers, adds them together and obtains the decimal 1354 in the binary representation:

$$0001 \times 1010^3 + 0011 \times 1010^2 + 0101 \times 1010^1 + 0100 \times 1010^0$$

1111101000	. . . .	1000
100101100	. . . .	300
110010	. . . .	50
100	. . . .	4
10101001010_2		1354_{10}

Here, $1010_2 = 10_{10}$.

This number is registered in the memory.

All subsequent operations are carried out by the computer automatically in response to pulses generated by the control unit in accordance with the program. The results of computation are read out in the reverse order, that is, binary numbers are converted into binary-coded decimals and these to decimal numbers, and printed by an electric typewriter.

Chapter Twenty Three — Industrial Applications of Electronics. An Outline of Automation

23-1. Automatic Systems

Automation in industry is a vital venue in the technical progress of economy. It gives a basis for a continuous rise in labour productivity and in the quantity and quality of output.

More specifically, automation is the field of science and engineering concerned with the design and use of automatic devices, mechanisms and machines—in short, all facilities that enable industrial processes to be run without direct intervention from man. Indeed, automation relieves man of arduous work and leaves for him only to set up, start and supervise the plant.

In Soviet usage, the word "automatic control" applies when controllers and the controlled plant are separated by a short distance. If the spacing is large, it is usual to speak of telecontrol.

A huge variety of automatic devices are used by industrial automation. We shall only deal with systems using electrical and electronic means. For the purpose of our discussion, we shall class them all into three broad groups, namely, automatic inspection and quality control, automatic machine control, and automatic process control.

23-2. Elements of an Automatic System

The elements, or components, that make up a typical automatic system may be classed into:

— measuring or sensing elements, more generally called transducers;

— relays and switches;
— amplifiers;
— actuators and servo motors, that is, final control elements.

Transducers measure the values of the physical or chemical quantities most representative of the controlled process or plant. In a more restricted meaning, the words "sensing element" are often applied to the first element or group of elements which respond qualitatively to the variable in question, and the word "transducer" to the device which generates a signal suitable for transmission to intermediate, indicating or control devices. Transducers have been examined in Sec. 8-8.

Relays and switches are intended to open and close measuring and control circuits. Some of them have been discussed in Chapters 11 and 21.

Amplifiers are intermediate devices intended to boost the signals generated by transducers to values sufficient to drive final control elements (actuators and servo motors). They have been examined in Chapter 19.

Final control elements (actuators and servo motors) cause the controlled variable to change in a predetermined manner.

23-3. Automatic Inspection and Quality Control

Automatic inspection and quality control covers the dimensions and quality of finished products, the concentration, turbidity and colour of solutions, the conditions (temperature, pressure and other variables) of a process, etc.

Also falling under this heading is the counting, sorting and rejecting of finished products.

If inspection is carried out at individual work stations, we have local or single-point inspection or quality control. The data thus gathered can be relayed to a central location so as to give an overall picture; this constitutes centralized or multipoint inspection or quality control.

In Soviet usage, if the distance between an inspection station and the central room is small, reference is made to remote inspection. If the distance is large and involves

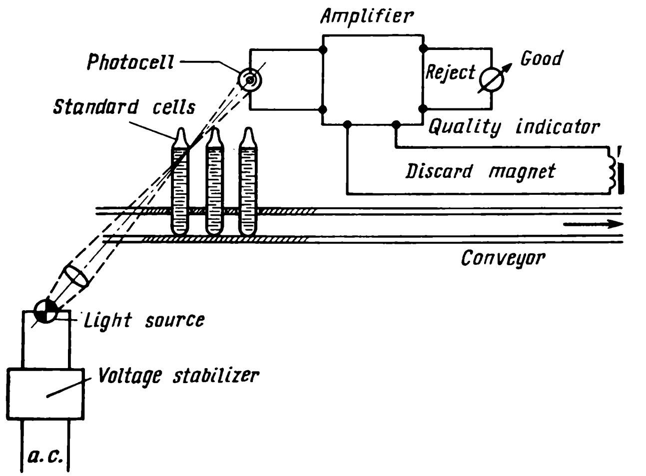


Fig. 23-2. Automatic inspection of liquids for turbidity

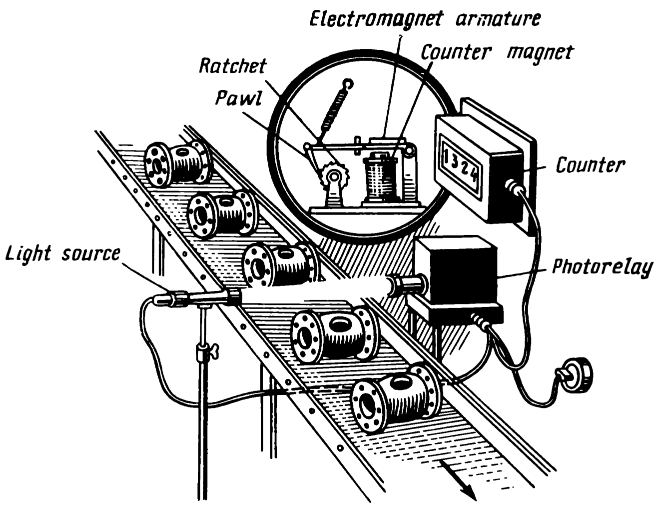


Fig. 23-3. Counting workpieces automatically

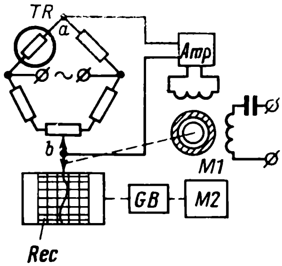


Fig. 23-4. Temperature logger

is boosted by an amplifier, *Amp*, which energizes the control winding of a reversible motor, *M1*. The other winding of the motor is energized by a.c. from the supply line.

The current in the control winding causes the rotor to turn and actuate, via a gear box, a wiper *b* which adjusts the resistances in the arms to a value at which the bridge is again at balance. The displacement of the wiper necessary to restore balance is a measure of temperature variations, and these are recorded by the stylus of a recorder, *Rec*, on a paper chart, thus producing a permanent record, or log. The chart drum is driven by a constant-speed synchronous motor, *M2*, via gear box, *GB*.

23-4. Automatic Machine Control

In this section we shall deal with automatic devices which cause the actuators and servo motors of the controlled plant to operate in such a manner that the operating variables are maintained or varied as may be required by the production process. This mainly covers speed control, braking, reversal, rotation of mechanism through an angle, advance of the workpiece, etc.

(a) Automatic Drive Control

Most often, electric drives are controlled automatically. As an example, let us see how a d.c. motor is started automatically.

A shunt-wound d.c. motor is started by applying supply voltage via a starting rheostat made up of a number of

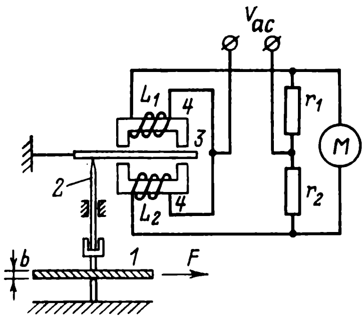


Fig. 23-1. **Circuit for automatic quality inspection**

the use of communication facilities, one has telemetry and supervisory indication.

Most commonly, the purposes listed above are served by electrical means, because they are applicable to all kinds of quantities, both electrical and nonelectrical, and can keep continuous watch on the variable of interest for monitoring and recording at any distance, and do so to a high level of accuracy and sensitivity.

Let us discuss several examples.

(1) In Sec. 8-8 we have discussed a variable-resistance transducer used for the remote indication of the level or volume of a liquid (see Fig. 8-35).

(2) Figure 23-1 shows an automatic inspection circuit using a variable inductance. The workpiece, *1*, under inspection is advanced in the direction marked by the arrow on a belt conveyor. Depending on its thickness, b, the sensing spindle, *2*, will move up or down, thereby varying the air gaps between the armatures, *3*, and cores, *4*, of an electromagnet coils. In this way, the inductance of the coil is made to vary in proportion to the thickness of the workpiece, *1*.

The thickness, b, of a workpiece can be measured by, say, an a.c. bridge similar to a d.c. bridge, with the coils L_1 and L_2 placed in one pair of bridge arms, and resistors r_1 and r_2 in the other pair. With the voltage across one pair of bridge junctions held at a constant value, that across the other pair will be a function of the thickness

of the workpiece. The meter can be calibrated to read units of thickness directly.

(3) Products are often inspected by means of photocells. For example, in measuring the diameter of a wire, the latter is made to move in front of a photocell. A heavier wire will intercept a greater proportion of the luminous flux incident on the photocell, and the photocurrent generated will decrease. The diameter can be indicated by a suitably calibrated current meter connected via an electronic amplifier to the photocell.

Figure 23-2 shows an arrangement for the automatic inspection of liquids for turbidity. Light from a lamp passes through a vial holding the liquid under inspection and reaches a photocell. The amount of luminous flux reaching the photocell will depend on the turbidity of the liquid, and so will the photocurrent measured by a meter (indicator). Should the turbidity of the liquid exceed a predetermined level, a discard electromagnet will remove the vial from the conveyor.

(4) Figure 23-3 illustrates the use of a photocell for counting workpieces. Light from a light source strikes a photocell and causes it to generate a photocurrent. A workpiece carried past the photocell by the conveyor intercepts the light beam, the photocurrent drops and causes a photorelay to operate. Each time the photorelay operates, a current pulse is sent to a counter. In the counter, the armature is attracted by the core and causes the pawl linked to it to advance its ratchet one tooth forward. The counter mechanism actuated by the ratchet has four drums with the digits from 0 to 9 inscribed on the side surface of each. The digits on the first drum read units, those on the second read tens, on the third, hundreds, and so on.

(5) Figure 23-4 illustrates the operation of an automatic temperature logger.

The temperature sensor is a thermistor*, TR, placed in one arm of a bridge. Initially, the bridge is at balance. As the temperature varies, the resistance of the thermistor changes, the balance of the bridge is upset, and a current begins to flow between bridge junctions a and b. This signal

* A thermistor is a temperature-sensitive resistor

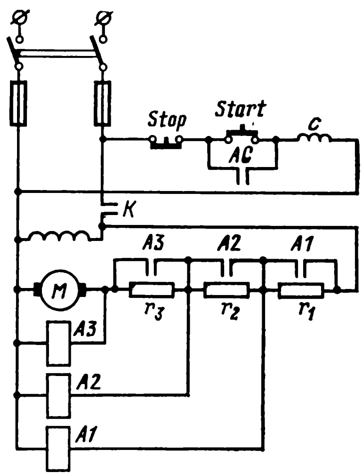


Fig. 23-5. Automatic starting of a d.c. motor

resistors (r_1, r_2, r_3) connected in series with the armature. Gradually, these resistors are switched out of circuit. With a low-power motor, this is done automatically, with an increase in motor speed or in its emf which is proportional to its rpm.

Figure 23-5 illustrates the automatic starting of a d.c. motor. After a two-pole knife-blade switch is cut in and the "Start" button is pressed, the contactor coil is energized and causes the main contacts, *K*, and the auxiliary contacts, *AC*, to make. The latter shunt the "Start" button and maintain the supply circuit to the contactor after the button is released. At the same time, a current begins to flow in the field winding, the motor armature, and the sections r_1, r_2 and r_3 of the starting rheostat connected in series with the armature. Now the motor begins to rotate, picks up speed, and its emf rises in proportion. As speed rises, the first acceleration contactor, *A1*, operates and closes its auxiliary contacts which shunt the first section, r_1, of the starting rheostat. As speed keeps rising, the second acceleration contactor, *A2*, operates and closes its auxiliary contacts which shunt the second section, r_2, of the starting

rheostat. Finally, the third acceleration contactor, *A3*, operates and closes its auxiliary contacts which shunt the last section, r_3, of the starting rheostat. This completes the starting sequence. The motor can be stopped by pressing the "Stop" button.

(b) Time-Sequence Control of Furnace Temperature

The sequence in which furnace temperature is to be varied with time can be set up in advance as a suitably profiled cam, as is shown at *1* in Fig. 23-6. The system illustrated in this figure continually compares the actual temperature in the furnace, *9*, with one that must be at any particular time, as preset on the cam actuated from right to left by a clockwork, *2*.

The set-point value of temperature is sensed by a roller, *3*, riding the surface of the cam and relayed to a bellcrank, *4*, mounted on a pivot, *5*. The pointer, *6*, indicating the furnace temperature is placed in a vertical slot in the bellcrank and linked to a measuring potentiometer, *7*. If the actual furnace temperature differs from its set-point value, the pointer will touch one of the contacts on the bellcrank and complete the supply circuit to a servo motor, *8*. The motor starts rotating and varies the feed of fuel to the furnace so as to minimize the temperature error, as the difference is usually called. When the actual temperature is the same as its set-point value, the pointer does not touch any of the contacts on the bellcrank, and the servo motor remains stationary.

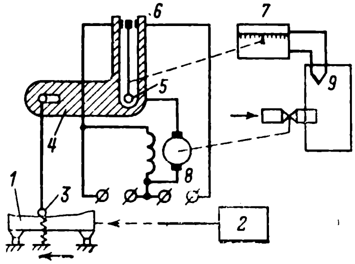


Fig. 23-6. Time-sequence control of furnace temperature

(c) Telecontrol on Railways

The safety of traffic on railways is ensured by what is known as an automatic block signal system. Such a system is an assemblage of automatic and telecontrol facilities intended to time railway traffic in such a way that a train will be permitted to occupy the advance block (segment of track) by the traffic light at the start of the block only after the previous train has left it.

In fact, the train itself actuates the associated signalling devices via the track circuits. The traffic circuit of a block is isolated from those of the blocks ahead and behind by insulated rail joints.

On electric-traction railways, the track circuits operate on alternating current at 25 or 75 Hz, if the frequency of traction current is 50 Hz. If traction uses direct current, the track circuits operate at 50 Hz.

The continuity of circuit for traction current is ensured by track choke-transformers (Fig. 23-7) which present a physical break to the track-circuit current between the separate blocks. One of the functions of the track circuits is to transmit the signals to be displayed by the traffic lights in the adjacent blocks. Here signals are transmitted as combinations of current pulses. For example the green light can be represented by three current pulses and a long interval separating them from the next three pulses; the yellow light can be represented by two pulses, and the red light by one pulse. As is seen, this is a number code, and the system is quite appropriately called a code signal system.

Pulse combinations are sent into the track circuit by a code transmitter situated at the end of each block (see Fig. 23-7). At the start of the next block these pulses are received by a track relay at the traffic light, and the relay operates in synchronism with the incoming pulses. The pulses are then converted by a decoder in such a way that the traffic light is caused to display an appropriate colour. Arrival of three pulses causes the decoder to turn on the green light. Two pulses will cause it to turn on the green light, too, because the advance block is clear, but its traffic light displays yellow colour because its track relay has

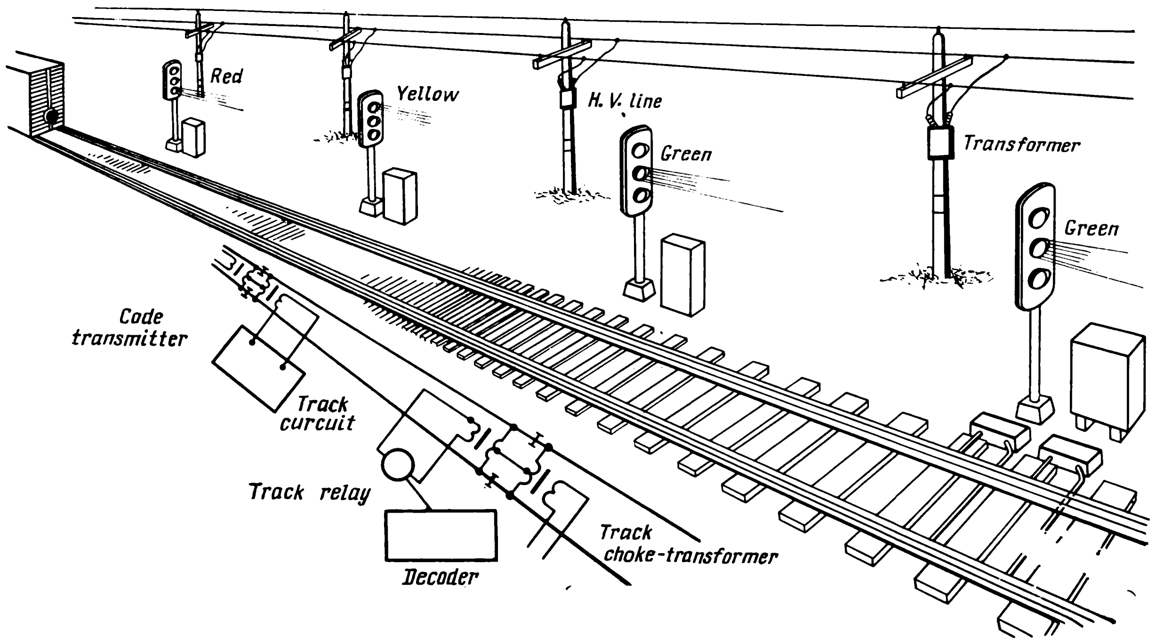


Fig. 23-7. Automatic block signal system of a railway

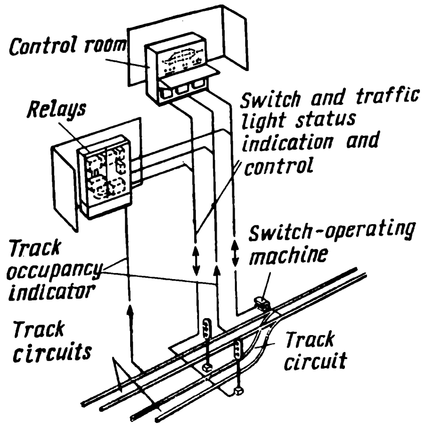


Fig. 23-8. Some of the devices used in an automatic railway signal system

received a single pulse from the occupied block. No pulses are fed to the traffic light of the occupied block, and it displays red colour. As soon as the train clears the block, one-pulse signals will arrive at its traffic light, and the latter will display yellow colour, thereby causing two pulses to be sent into the track circuit, so the light traffic in the previous block will display green instead of yellow colour. These code combinations may be used for automatic block and train signalling.

The automatic block signal system takes its power from the secondary of a transformer at 110-120 V fed over a cable to the relay cabinet. The transformer primary is usually connected to a power transmission line operating at 6-10 kV.

On present-day railways, switches and signals are usually controlled from a central control room. The devices used for the purpose constitute a centralized switch and signal control system. With it, control of a large number of switches and signals can be concentrated in a single control room. It speeds up route selection, enhances traffic safety, and raises the efficiency of personnel.

A typical centralized switch and signal control room has a control board, an illuminated track diagram, control devices, track-circuit interlocks, and power supplies (Fig. 23-8). The switches and traffic lights are situated

at their locations throughout the railway station, and the switches are actuated by their operating machines. The switches, traffic lights and other track devices are connected to the control room by cables.

Signal and switch interlocking is only one of the functions performed by a centralized control room. It also must prevent admission of a train to an occupied track and transfer of switches under a train; to this end, it must keep continuous watch on the status of switches and the occupancy of station tracks.

For control of switches and signals, the control room has control panels or boards. At small stations, each switch is ordinarily operated separately from any other, by means of two buttons. The position occupied by a switch at the moment is indicated by an illuminated lamp on the control panel. After the switches have been placed in the position corresponding to the selected route, the operator presses a suitable button, and the outgoing traffic light displays green colour.

The route is selected by consecutively pressing the two route buttons on the control panel, one button marking the start and the other button the end of the route (this is known as route control).

The operator can use his control panel only after he receives an acknowledgement that all the devices have executed the respective commands, and after he ascertains the status of all controlled switches, traffic lights, tracks and switch stands. Their status is monitored by reference to the illuminated track diagram which has lamps to indicate the occupancy of the inward- and outward-bound tracks and switch sections by trains. When a track or a track section is not occupied, the respective lamp on the track diagram remains dead.

A system combining automatic block signalling, electric interlocking and control of all switches and lights at the intermediate stations of a track section by a single dispatcher from his room constitutes a centralized dispatching (or centralized traffic) control system.

In this case, train handling at a way (intermediate) station reduces to route selection, operation of traffic lights, and dispatch of trains in the sequence ordered by the dis-

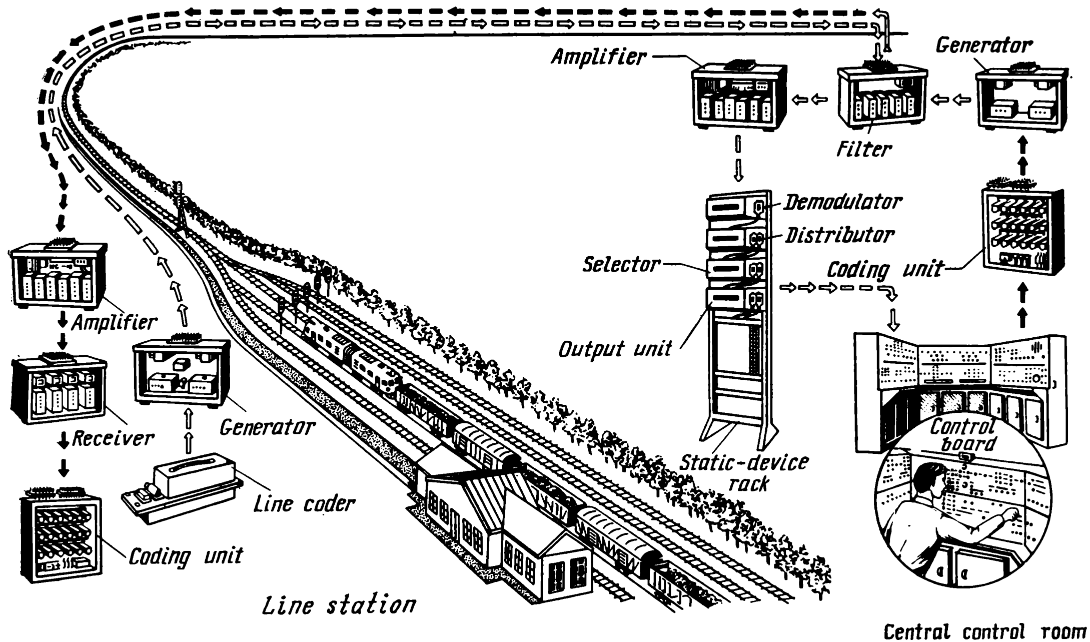


Fig. 23-9. General layout of a centralized dispatching control system

patcher. The facilities available enable one train dispatcher to control all the crossing loops and intermediate stations within a single track section (100 to 150 km long) (Fig. 23-9).

The capabilities of centralized dispatching control include control of all switches and station signals within the section from a single point, monitoring the status of switches, traffic lights[1] and isolated sections by means of supervisory devices, and recording of train traffic by a train recorder.

Centralized dispatching control uses a code control system which transmits commands from the train dispatcher to the outlying stations (telecontrol) and receives messages from the outlying stations to the train dispatcher about the status of controlled switches and lights, the occupancy of tracks and switch sections, the position and direction of travel of trains (teleindication).

The equipment that the train dispatcher has at his disposal in the central control room includes a control board, an illuminated track indicator, a train recorder, a central encoder to transmit commands to and receive messages from the outlying stations within the section. Commands are sent from the control board as the dispatcher presses the respective command buttons. Where route selection involves the selection of the respective traffic lights, the dispatcher first uses the button corresponding to the start of the route (the "inward-bound" button), then the button corresponding to the end of the route (the "outward-bound" button). In this way, he assigns a track for the train and the traffic light that will give a "clear" signal to the train's driver.

(d) Numerical Control of Machines

Advances in computational engineering have served much to stimulate the development of numerically controlled machines.

Basically, numerical control of machines consists in working out a sequence of operations required to make a part, one also compiles a table (called a program sheet) showing the discrete points (in position control) or continuous path (in path control) that the tool will take up or

follow in performing its function. These positions or the continuous path are encoded in numerical form, that is, as a set of numbers. The numbers are then punched into a card or a tape, and the card or tape is finally loaded into the punched-card or punched-tape reader of the machine. The reader converts the combinations of punchings into appropriate commands which are distributed by suitable devices among the final control elements (slide actuators, indexing-table actuators, etc.).

All numerical control systems mainly differ in the manner in which position or path data are generated and converted and also the manner in which the commands are executed by the final control elements. Accordingly, there are open-loop, closed-loop and self-adjusting numerical control systems.

Figure 23-10 shows an open-loop numerical control system. A program stored on punched tape, *2*, is loaded into a punched-tape reader, *1*, which converts the combinations of punchings into respective commands in the form of electric signals. The signals undergo one more conversion in block, *3*, whence they are routed to the actuator, *4*. As a result, the actuator causes, say, the slide or indexing table to move to the next position prescribed in the program. Whether or not the actual displacement is the same as the one prescribed in the program is not checked.

Figure 23-11 shows a closed-loop (or feedback) numerical control system. In this system, feed-forward data read off the storage medium, *2*, by a reader, *1*, are routed via converter, *3*, to a comparator, *4*. Feedback data are generated by a feedback sensor, *7*, which monitors the actual displacement or position of the tool and directs the signal likewise to the comparator. From a comparison of feed-forward and feedback data, the comparator generates what is called an error signal which drives an actuator, *5*, and this moves the tool, *6*, as prescribed by the program and conveyed by control signals. As soon as the actual displacement (or position) becomes the same as the programmed one, the comparator ceases generating error signals.

In contrast to a closed-loop system, a self-adjusting system has an additional transducer which generates data about the actual characteristics of the surface being machi-

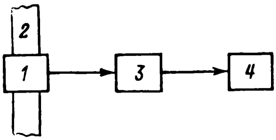


Fig. 23-10. Open-loop time-sequence control system

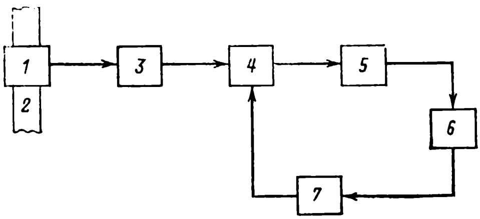


Fig. 23-11. Closed-loop (feedback) time-sequence control system

ned, and directs them to an interpolator. The latter combines the incoming data with those from the program reader and stores the result in internal memory for further use.

As we have learned, numerical control uses numbers in its programs (see Chapter 22). In position control systems, these numbers define the position or displacement of the slide. In path control systems, the numbers usually define the contour of the surface being machined relative to some reference point.

As an example, let us take a closer look at the numerical control of a vertical milling machine. The blank set up on the table of the machine is milled to produce a cam having the desired profile by causing it to move relative to a rotating cutter whose axis remains in one and the same position at all times. The blank is in two motions at a time, namely rotation, S, about its axis (Fig. 23-12) and translational motion towards and away from the cutter axis.

Programming involves a number of steps, such as machine set-up, determination of interpolation intervals, location of points on the workpiece, location of breakpoints in the tool centre path at equal increments, determination of the first difference between the locations of the cutter centre path breakpoints, and division of the result by the

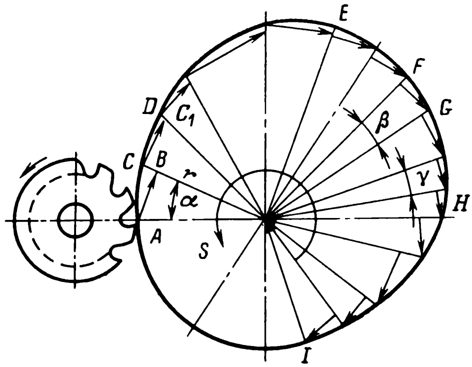


Fig. 23-12. Generation of a cam profile

elementary pitch or, which is the same, the incremental pulse. The latter is selected according to the surface finish desired and the tolerances on the work profile. For each angular unit of blank rotation, the pulses should take an appropriate sign, according to whether the cutter should move towards or away from the workpiece. For linear interpolation, the difference between adjacent values of radius r (see Fig. 23-12) is determined, that is, at points A and C, C and D, etc. After the blank is turned through an angle α, the cutter would move from point A to point B instead of point C. To avoid this, the cutter is fed from point B to point C, etc. No feed is applied between points H and A.

After input data have thus been determined they are transferred by a photographic camera onto photographic film used as the storage medium. The first track on the film is a record of programmed variations in the path of the feed servo. Light bars on the film stand for commands that cause the blank to move away from the cutter, and dark bars to move towards the cutter. The second track is a record of the programmed rate at which feed is to be varied. The third track stores the feed control program.

Operation of a numerically controlled vertical milling machine is illustrated in Fig. 23-13. The table mounting the blank, *14*, is driven by a motor, *3*. Longitudinally, the table is actuated by another motor, *10*, via a gear train,

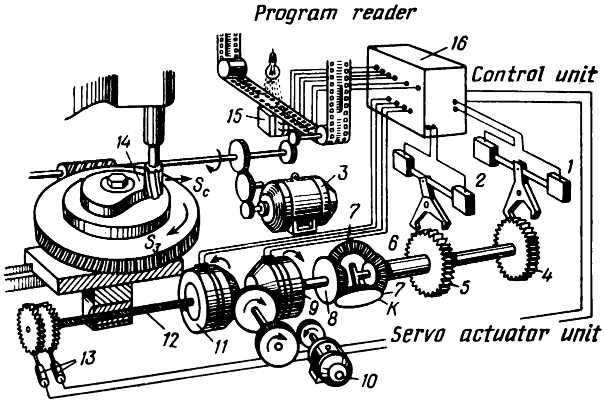


Fig. 23-13. Sketch of a numerically controlled vertical milling machine

electromagnetic clutches, *9* and *11*, and a lead-screw, *12*. The blank moves away from or towards the cutter, according to which one of the two electromagnetic clutches is energized. When control pulses recovered from code combinations on the film are applied to electromagnets, *1* and *2*, the associated escape pallets swing to and fro between the teeth of the escape wheels, *4* and *5*, the former having fine teeth and the latter coarse teeth. As a result, the lead-screw advances through a distance corresponding to the angle between two adjacent teeth on the escape wheels. In the absence of control pulses, the pallets and escape wheels remain stationary, thereby locking the lead-screw and the respective clutch slips.

The amount of longitudinal feed applied to the table is controlled by the tooth pitch of the gears in the gear train. Which gear is engaged depends on which relay is caused to operate. Operation of a relay is initiated by a pulse generated by a photocell, *15*, as it reads code combinations on the moving film, boosted by an amplifier, *16*, and applied to the actuator unit.

The lead-screw is linked to the escape wheels by a differential, *K*, and gears *6*, *7* and *8*. Feedback is effected as follows. At the left-hand end of the leadscrew, there are two toothed discs having the same number of teeth as the

escape wheels; these discs are coupled to feedback sensors, *13*. Should the lead-screw fail to rotate despite the application of a control pulse, the feedback sensors will generate an appropriate signal for the control unit, and the latter will cause the escapement to operate and the lead-screw to rotate and advance the workpiece as may be necessary.

23-5. Automatic Process Control

An automatic control system is intended either to maintain the progress of a process or plant to satisfy some predetermined criteria, or to vary this progress with time according to some law.

Such a process or plant is called the controlled process or plant, and the physical quantity associated with it and maintained or varied in a predetermined manner is called the controlled variable. The controlled plant may be, say, an electric motor, and the controlled variable may be its rpm.

The normal progress of a process or operation of a plant is subject to external influences, called disturbances.

A good many automatic control systems may graphically be depicted by a general block diagram such as shown in Fig. 23-14. Here, the block numbered "1" represents a transducer (measuring element or sensor) which senses variations in the controlled variable of the controlled plant or process, *CP*, which may be caused by external disturbances, *D*. The sensed value of the controlled variable is fed to a comparator, *2*, which compares it with the set-point value, *6*. If the two differ, the comparator generates an error signal which is boosted by an amplifier, *3*, to the desired value. A converter, *4*, converts the boosted signal to another form convenient for application to an actuator or final control element, *5*, and this acts upon the controlled plant (or process) so as to minimize the error signal.

Of course, in each particular case, a specific control system may have a different number of units, and the units themselves may perform functions differing from those discussed. Depending on the manner in which control systems affect the controlled variable, they may be classed into:

(1) control systems of the regulator type (this type inc

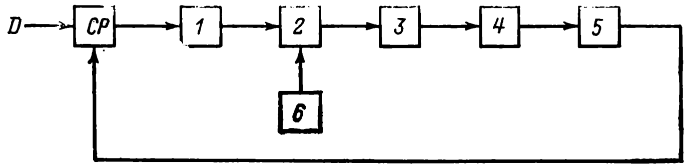


Fig. 23-14. Block diagram of an automatic closed-loop (feed-back) **control system**

ludes speed governors, voltage regulators, frequency regulators, etc.);

(2) control systems of the time-sequence type which maintain or vary the controlled variable to a predetermined time sequence (this type includes the time-sequence furnace temperature regulator examined in Sec. 23-4);

(3) control systems of the servo type in which a system responds to changes in the set point, these changes being of an unpredetermined nature.

Let us consider several examples.

(a) A regulator-type control system. Figure 23-15 shows the circuit of a d.c. generator voltage regulator. Any change in the load, R_L (external disturbance), of the generator, G, brings about a change in the generator terminal voltage, V_g. The difference between the set-point value, V_{set}, and the actual terminal voltage, V_g, that is, $\Delta V = V_{set} - V_g$, is applied to the input of an amplifier, *Amp*, whose output is coupled to the armature of a servo motor, M. Depending on the sign of the difference voltage, ΔV, applied to the amplifier, the armature will rotate clockwise or anticlockwise. In doing so, it will move the wiper of a rheostat connected in the field winding, *GFW*, of the generator. The servo motor will keep rotating and the wiper moved until the difference voltage, or error signal, is minimized to $V_g = V_{set}$. When this happens, the voltage applied to the amplifier will be zero ($\Delta V = 0$), the armature winding of the servo motor will be de-energized, and the motor will stop.

(b) A servo-type control system. Figure 23-16 shows a servo system used to transmit the angular position of an antenna to an output shaft. The antenna shaft is coupled to the wiper of a potentiometer, R_1, so the wiper setting

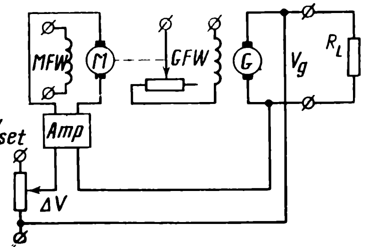


Fig. 23-15. Voltage stabilization of a d.c. motor

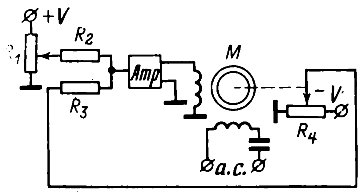


Fig. 23-16. Transmission of antenna angular position to an output shaft by a servo system

is proportional to the angular position of the antenna. Accordingly, the voltage, V_1, taken from the potentiometer wiper is likewise proportional to the angular position of the antenna. This voltage is applied via a resistor, R_2, to an amplifier, *Amp*, whose output is coupled to the control winding of a servo motor, *M*. The shaft of the servo motor is coupled via a gear box to the wiper of potentiometer, R_4.

The voltage, V_4, taken from the wiper of R_4 is applied via R_3 ($R_3 = R_2$) to the same amplifier. Because V_1 and V_4 are of opposite polarity, the voltage applied to the input of the amplifier will be their difference, $\Delta V = V_1 - V_4$. The servo motor and, as a consequence, the output shaft will keep rotating until the difference voltage, or the error signal, is minimized to zero. When this happens, the wipers of R_1 and R_4 will have taken up identical angular positions. In this way, the angular position of the output shaft serves as an indication of the angular position taken up by the antenna.

The requirements that any control system must above all satisfy are stability (freedom from hunting), accuracy, and high speed of response. To meet these requirements, every system incorporates suitable compensating elements.

Recently, direct digital control (DDC) systems have gained popularity. In them, a single computer controls different variables over a large number of control loops. Control action, or law, may be any, however complex, and can readily be adapted to a varying situation.

Index

Bibliography

1. BULANOV YU. A., GLAGOLEV G.I. Basic Electronics (*Osnovy electrotekhniki*). Moscow, "Higher School" Publishers, 1966 (in Russian).
2. BULANOV YU. A., GOLUBKOV G.N. Handbook of Laboratory Exercises in Basic Electronics. (*Rukovodstvo k provedeniyu laboratornykh rabot po osnovam electrotekhniki*). Moscow, "Higher School" Publishers, 1966 (in Russian).
3. ZHEREBTSOV I.P. Electronics. English translation by "Mir" Publishers, Moscow, 1970.
4. PETROV V.K. SHLIAPINTOKH L.S. Problems in Electricity and Basic Industrial Electronics (*Sbornik zadach po elektrotekhnike i osnovam promyshlennoy elektroniki*). Moscow, "Higher School" Publishers, 1965 (in Russian).
5. POPOV V.S., NIKOLAYEV S.A. Basic Electricity and Electronics (*Obshchaya elektrotekhnika s osnovami elektroniki*). Moscow, "Energia" Publishers, 1972 (in Russian).
6. POPOV V.S., NIKOLAYEV S.A. Electrical Engineering (*Elektrotekhnika*). Moscow, "Energia" Publishers, 1968 (in Russian).
7. RABINOVICH E.A. Problems and Exercises in Basic Electricity (*Sbornik zadach i uprazhneniy po obshchey elektrotekhnike*). Moscow, "Higher School" Publishers, 1972 (in Russian).
8. RABINOVICH E.A. Handbook of Laboratory Exercises in Basic Electricity (*Rukovodstvo k provedeniyu laboratornykh rabot po obshchey elektrotekhnike*). Moscow, "Higher School" Publishers, 1966 (in Russian).

TO THE READER

Mir Publishers welcome your comments on the content, translation, and design of the book.
We would also be pleased to receive any suggestions you care to make about our future publications.
Our address is:
USSR, 129820, Moscow, I-110, GSP, Pervy Rizhsky Pereulok, 2, Mir Publishers.

Printed in the Union of Soviet Socialist Republics

OTHER BOOKS FOR YOUR LIBRARY

Photometric Measurements in Television

by *M. Krivosheev*, D. Sc. and *A. Kustarev*, Cand. Sc.

A book for a wide circle of specialists engaged in the development and application of television transmitting and receiving equipment. Centres attention on the examination of methods and instruments for measuring the light parameters of objects and television images, as well as on the measurements in the light-signal and signal-light conversion. Also, outlines the use of television equipment in various branches of science and technology.

Electrical Network Performance Calculations and Analysis

by *V. Venikov*, D. Sc.

The textbook is meant for the students of electric power engineering institutes. Presents collected problems and examples of electrical network performance analysis and calculations. All the examples are supplemented with detailed solutions, illustrating curves, tables and explanations that will assist in a more comprehensive understanding of the theory of respective subjects.